The Applicability of a Machine Learning Methodology to Generate TMY Weather Files

Ashleigh Marie Papakyriakou

A Thesis

in

The Department

Of

Building, Civil, and Environmental Engineering

Presented in Partial Fulfillment of the Requirements

for the Degree of Master of Building Engineering at

Concordia University

Montreal, Quebec, Canada

April 2024

CONCORDIA UNIVERSITY

School of Graduate Studies

This is to certify that the thesis prepared

By: Ashleigh Papakyriakou

Entitled: The Applicability of a Machine Learning Methodology to Generate TMY Weather Files

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. Radu Zmeureanu

______________________________Examiner
Dr. Radu Zmeureanu

______________________________Examiner
Dr. Alireza Nazemi

______________________________Supervisor
Dr. Bruno Lee

Approved by ____
Chunjiang An, Graduate Program Director

____________________2024

____
Dr. Mourad Debbabi, Dean of Faculty

Abstract

The Applicability of a Machine Learning Methodology to Generate TMY Weather Files

Ashleigh Papakyriakou

To effectively decarbonize buildings accurate energy models must be created to predict building energy performance. Typical meteorological year (TMY) weather files represent long-term weather conditions and are used in energy modelling to help evaluate energy performance. This thesis explores generating TMY files using machine learning to improve accuracy, which can significantly influence energy simulation results. The current TMY generation approach relies on expert judgment, often overlooking seasonal, climate and application-based variations.

Manuscript #1 introduces a machine learning methodology using feature importance to determine the relevant generation parameters used in the Sandia method to enhance the current TMY generation approach. The proposed methodology is applied to a medium office building in Montreal. The results reveal an improved representativeness of the long-term average building energy demand for the TMY generated using the proposed methodology.

Manuscript #2 aims to (1) assess the applicability of the methodology across Canadian climates; (2) investigate the feasibility of using standardized climate zone-based weighting factors to reduce the computational time associated with extracting location-based weighting factors to facilitate wider adoption of the proposed methodology. The methodology is applied to 18 cities across six Canadian climate zones and generates two weather files for each location. $TMY_{STATION}$ uses location-based weighting factors while TMY_{CZ} uses climate zone-based weighting factors. The CV(RMSE) and NMBE indicate the proposed weather files outperform the conventional weather files in predicting the long-term energy performance of buildings. Although the $TMY_{STATION}$ files performed marginally better, the convenience of standardized climate zone-based weighting factors can enhance the methodology's adaptability.

Acknowledgements

I would like to express my gratitude and appreciation to my supervisor Dr. Bruno Lee for his assistance and support. A special thank you goes to my dear friend, Ana, whose valuable insights, contributions, and dedicated time played a pivotal role in finalizing this contribution. Your friendship and support have meant the world to me. To my partner, Jared, and my close friends, and family, I am thankful for your support and patience during this challenging journey, and for encouraging me not to give up.

In loving memory of my grandfather, John Papakyriakou, I dedicate this thesis. His enduring encouragement, guidance, kindness, and unwavering support has shaped me into the person I am today.

Table of Contents

List of Figures

List of Tables

Abbreviations

CWEC	Canadian Weather Year for Energy Calculation
CWEEDS	Canadian Weather Energy and Engineering Datasets
CV(RMSE)	Coefficient of Variation of the Root Mean Square Error
CZ	Climate Zone
DBT	Dry Bulb Temperature
DHI	Diffuse Horizontal Irradiance
DNI	Direct Normal Irradianece
HDD	Heating Degree Day
IWEC	International Weather for Energy Calculation
LTA	Long-Term Average
LTW	Long-Term Weather
MPE	Mean Percentage Error
NECB	National Energy Code of Canada for Buildings
NMBE	Normalized Mean Bias Error
RH	Relative Humidity
RMSE	Root Mean Square Error
SEE	Standard Error of Estimates
TMY	Typical Meteorological Year
WS	Wind Speed
WWR	Window Wall Ratio

Chapter 1 Introduction

1.1 Background

The buildings sector is the third-largest source of greenhouse gas (GHG) emissions in Canada [1]. Canada has an ambitious goal of reaching net-zero emissions by 2050 [1]. Achieving this goal involves the decarbonization of both existing and new buildings. Energy modelling is an important tool used to evaluate the energy consumption of various design alternatives and is frequently used to assess design alternatives in buildings. Energy modelling aids in selecting the most cost-effective and impactful design decisions, thereby reducing a building's energy consumption, carbon emissions, and operating costs.

Designers and engineers typically use a typical meteorological year (TMY) weather file in energy modeling, rather than multiple years of historical weather data, to facilitate comparison and evaluate different designs. A TMY weather file is a synthetic file created for a specific location to represent typical long-term weather conditions. The TMY weather file is composed of 12 typical months of hourly historical weather data and does not include extreme weather data. Typically, a minimum period of ten years is analyzed to create these files.

The Sandia method [2] [3] is a common method used to generate TMY weather files, other well-known TMY generation methods are the Danish method [4] and the Festa and Ratto method [5], however, these generation methods can be more complex to apply [6]. The Sandia method uses the Finkelstein-Schafer (FS) statistical method and weighting factors to select individual months from long-term historical weather data that represent the most typical month for a given location. The weighted sum is determined for each month and the month with the lowest weighted sum is selected. The complete Sandia method is outlined in Chapter 2.

Provinces across Canada such as Ontario [7] and British Columbia [8], as well as municipalities such as the City of Toronto [9], have sustainability requirements that must be met, which are typically verified using an energy model with a TMY weather file such as the Canadian Weather Year for Energy Calculation (CWEC) [10]. The CWEC TMY weather file datasets were developed by Environment Canada and are commonly used to represent Canadian locations. Many other commonly used TMY weather files use universal weighting factors which were assigned based on the intended use of the weather file using expert judgement. The original Typical Metrological Year ([TMY][1]) [11], dataset was designed to be used for solar heating systems [2] while TMY2 [12] and TMY3 [13] were designed to be used for energy conversion and building systems [12]. Furthermore, the CWEC, International Weather for Energy Calculation (IWEC) [14], and IWEC2 [14] [15], weighting factors were selected to best represent the building systems [10].

The conventional practice of assigning universal weather parameters and annual weighting factors based on expert judgement is inadequate in capturing the diverse requirements of different applications and the variation in the climatic characteristics of a location. Given that TMY weather files exert a significant influence on the simulation results, it is imperative to use TMY weather files that accurately represent long-term weather conditions.

1.2 Literature Review

1.2.1 <u>Impact of Weather Data on Results</u>

Weather has a significant influence on building energy consumption. Using a TMY weather file produced with poor data quality or TMY weather files that do not accurately represent a location can skew simulation results. A study by Bhandari et al. [16] compared two sources of third-party historical weather data from different providers with measured data (referred to as Meas) for the study location, Oak Ridge Tennessee in the US. The authors observed significant variation between the data sets, with differences in individual hourly reaching as high as 90%. This variation can be attributed to the raw data and processing techniques used to produce the weather parameters. Additionally, the authors used building simulation to compare the three data sources with a TMY3 weather file for Oak Ridge. The study found significant variation in heating and cooling loads, with monthly loads varying by up to 40%, however, the annual energy consumption only varied by up to 7%. The study highlights the importance of the quality of the historical data used to generate TMY weather files and the influence using poor data can have on simulation results.

[1] *[TMY]: Used to denote the Typical Meteorological Year file format originally presented in* [2]

Furthermore, the study highlights, how weather files produced by different organizations can yield different results.

There have been studies completed to determine if TMY weather files can accurately reflect the annual fluctuations in long-term weather data. A study by Hong et al. [17] compared simulation results for 30 years of historical weather data with TMY3 weather file results for three different sizes of prototypical office buildings (small, medium, large) at two different design efficiency levels across 17 ASHRAE climate zones. The study concluded with four main findings: 1. The variation in annual weather data has a larger impact on peak demand compared to building energy consumption; 2. Building simulations completed with TMY3 weather files can significantly under- or overestimate energy consumption and do not provide a good representative of the average energy use using AMY data across a 30-year period; 3. Buildings in colder climates tend to be more sensitive to annual variations in weather data; 4. The medium office building was most sensitive to variations in annual weather data compared to large and small offices. The study highlights the need to improve TMY weather files to better represent the long-term average data. Furthermore, the study demonstrates how different locations and different building types can have varying sensitivity to weather data. A study by Cui et al. [18] compared TMY weather files with historical long-term weather data for major cities in China to determine if TMY weather files can represent the annual seasonal variation. The study concluded 1. Colder climates have more significant variation in annual weather data, 2. The TMY weather file provided a good representation in terms of the long-term average but did not represent the variation in weather data 3. TMY weather files tended to over and under-estimate the peak load and energy consumption 4. The peak demand was more sensitive to variations in weather data than energy consumption. The study advocates for TMY weather files to be generated with customized weighting factors that account for the individual climate characteristics of a given location.

Lastly, a few studies compared the impact on results of using the different TMY generation methods to create TMY weather files. A study by Janjai and Deeyai [6]. compared the Sandia method, the Danish method, and the Festa and Ratto method to generate TMY weather files in Thailand. The study found the three methods produced similar results; however, the authors recommended the use of the Sandia method due to its simplicity in application. However, a study by Skeiker [19], compared the three methods by generating TMY weather files with 10 years of hourly data for Damascus, Syria, and found the Sandia method best represented the long-term average performance. These differences in results between the two studies may be attributed to the different climates between the locations. Thailand is characterized by a year-round tropical climate [6] whereas Syria is characterized by a Mediterranean climate and experiences very dry summers and mild winters [20].

The studies summarized above emphasize the following:

- Weather data has a significant impact on building simulation results.
- The data quality, and how organizations process the data can impact the results.
- The current TMY weather files may not adequately reflect the fluctuations seen in the long-term weather data.
- Not all TMY weather files perform the same.

Based on the following points, there is a need to improve TMY weather files to reflect long-term weather data more accurately.

1.2.2 TMY Weather File for Different Applications

TMY weather files can be used for various applications such as building energy simulation, renewable energy systems, daylighting, hydrology, hygrothermal, climatology, meteorology, and biometeorology. However, the weather parameters and weighting factors must be determined based on the application of the TMY weather file, as stated in the original [TMY][1] weather file manual [2].

Georgiou et al. [21] conducted a parametric analysis in Cyprus, assessing the influence of weighting factors on TMY generation for various applications, including residential solar thermal systems, wind turbines, and typical dwellings in Cyprus. The study demonstrated the influence the weighting factors have on the typical months selected to create the TMY weather file.

Several studies have modified the weighting factors used to generate TMY weather files based on the application using the Sandia method and saw an improved representation with the long-term average compared to the typical weighting factors used. Kambezidis et al [22] created TMY weather files for five different applications which were meteorology-climatology, biometeorology, agro-meteorology-hydrology, photovoltaics, and building energy models. They found the generated TMY weather files to provide a good representation of the long-term average however, the authors selected the weights based on judgment for each application.

Two studies used a genetic algorithm to optimize the weighting factors used to generate TMY weather files based on the intended use of the weather files. A study by Sun et al. [23] used a genetic algorithm (GA) to optimize weighting factors for a TMY weather file to be used for both daylighting and energy simulation. Chan [16] also used a GA to optimize weighting factors to create TMY weather files for each of the following applications: a fully air-conditioned office building, a building-attached non-concentrating photovoltaic (BaPV) system, a wind turbine power generation system, and a concentrating solar power (CSP) system. In the study, Chan [16] found the optimized weights for each application to vary compared to the original IWEC weather file weights, however, found the weights for the office buildings to be closest to the IWEC weights.

The above studies indicate the importance of determining weightings for the TMY weather files based on the intended use of the weather file. However, in the above studies, the weather parameters are manually assigned, and the weightings are either manually assigned or optimized using GA. A drawback of using GA to optimize weighting factors is that it can be time-consuming to determine the optimized weights.

1.2.3 TMY Weather Files for Different Climates

Many studies have evaluated the impact of TMY weather files on different climates. A study by Kalamees et al. [24] conducted a sensitivity analysis on different weather parameters and assigned monthly weights based on the amount of influence each parameter had on the heating and cooling demand. The study was conducted on an office building and a detached house in a cold climate. The study highlighted the seasonal variation in the influence of the weather parameters, demonstrating the need for monthly weighting factors.

In a study by Meng et al. [25], office buildings were simulated in China's three major climate zones to determine which climatic variables had the most significant impact on heating loads. The study found dry bulb temperature to have the largest influence on heating demand across the three climate zones however the magnitude of influence varied based on the climate zone as well as the influence from the other weather parameters. The study highlighted how the influence of weather parameters varies based on the climate zone.

A study by Qian et al. [26] proposed a methodology to generate new weighting factors for TMY weather files. The study was conducted on a multi-family low-rise apartment building for three cities in China with distinct climate zones: a severe cold climate zone, a cold climate zone, and a hot summer and cold winter climate zone. The TMY weather files for each location were generated using customized weighting factors determined using correlation. The study simulated the building energy model with the long-term weather data and compared the correlation using the normalized Pearson correlation coefficient between the simulated indoor temperature and the meteorological weather parameters to determine the weighting factors. Once the weights were determined, the study created TMY weather files using the Sandia method for each location and compared them with the original TMY datasets. The study found that TMY weather files generated using customized weighting factors had a better fit to the long-term average compared to the original TMY dataset and improved thermal comfort. Furthermore, the study found the correlation between indoor temperature and both wind speed and relative humidity varied based on the climate zone, further indicating the need for climate-specific weighting factors. One limitation of the proposed methodology is the use of the Pearson correlation coefficient to determine weighting factors. The methodology assumes a linear relationship between indoor temperature and each weather parameter. As a result, correlation may not be suitable to determine weighting factors for certain weather parameters. Additionally, the study is only evaluating correlation and does not consider multi-collinearity.

A study by Li et al. [27] used a new method to generate TMY weather files. The authors applied an entropy-based Technique for Order Preference by Similarity to Ideal Solution (TOPSIS) method to generate unique weighting factors and TMY files for different climate zones in China. The authors found the results to align better with the long-term average compared to using the Sandia method with fixed weighting factors. However, the study did not highlight whether the improvement was a result of the proposed generation method or the use of customized weighting factors. Furthermore, although a novel approach, the TOPSIS method is more complex than the commonly used Sandia method.

The following studies demonstrate the need for weighting factors to consider location-specific climatic characteristics however there still lacks a sufficient methodology to develop these weighting factors.

1.3 Objective

The objective of this thesis is to:

1. Determine if integrating machine learning into the generation of typical meteorological year (TMY) weather files improves the representativeness of how building energy models will perform over the long term compared to the current TMY generation approach.
2. Determine if the machine learning methodology is applicable to various climates.
3. Determine if the use of standardized climate zone-based weighting factors can enhance the adaptability of the proposed machine learning approach.

1.4 Thesis Structure

This manuscript-based thesis consists of four chapters:

- Chapter 1 introduces the thesis background, literature review, thesis objectives, thesis structure, and the manuscript executive summary.
- Chapter 2 is the first manuscript titled "Defining generation parameters with an adaptable data-driven approach to construct typical meteorological year weather files".
- Chapter 3 is the second manuscript titled, "Evaluating the applicability of a machine learning methodology to improve TMY weather file generation for different Canadian climate zones".
- Chapter 4 consists of the conclusions and future works.

1.5 Manuscript Executive Summary

1.5.1 Manuscript 1: Defining generation parameters with an adaptable data-driven approach to construct typical meteorological year weather files.

Chapter 2 introduces a methodology to systematically define the relevant weather parameters and calculate their corresponding weighting factors using machine learning and feature importance. The weather parameters and weighting factors are then integrated into the Sandia method to generate the TMY weather file. The methodology aims to improve the current approach of TMY weather file generation by providing a data-driven approach that is adaptable to different applications and climate zones.

The methodology uses a gradient-boosted tree regression model. The use of machine learning to determine the relevant weather parameters and weighting factors tailored to specific applications and locations minimizes the uncertainty associated with assigning weighting factors based on experts' judgement. The methodology was applied to a prototypical medium office building located in Montreal, Quebec. 20 years of hourly historical weather data was obtained, and a building energy model was created based on the National Energy Code of Canada for Buildings 2020 requirements for Montreal's climate zone. The methodology determined the relevant weather parameters that had the most influence on building energy demand and a set of 12 monthly weighting factors were generated using feature importance. Monthly weighting factors were obtained instead of annual weighting factors to account for seasonal fluctuations. The relevant weather parameters and weighting factors were then integrated into the Sandia method to generate the TMY weather file.

The generated TMY weather file was evaluated against the CWEC weather file for the same location, and both were compared to the long-term average historical weather data. The root-mean-square error (RMSE) was used as the performance indicator and the generated TMY weather files had lower RMSE values indicating a better representation of the long-term average energy demand. The generated TMY weather file improved the performance by 16% compared to the CWEC weather file. Although the proposed methodology performed better than the CWEC weather file, the methodology still has a few limitations such as being location-dependent, the accuracy of the machine learning model, and requiring continuous data inputs.

1.5.2 Manuscript 2: Evaluating the applicability of a machine learning methodology to improve TMY weather file generation for different Canadian climate zones.

Chapter Chapter 3 applies the methodology proposed in Chapter 2 to 18 locations across Canada to represent the six diverse Canadian climate zones. The purpose of the study is to:

1. Determine the applicability of the proposed machine-learning methodology across various climates.
2. Determine if the use of a standardized set of climate zone-based weighting factors is feasible to enhance the adaptability of the proposed methodology.

The study produces two TMY weather files for each location, a $TMY_{STATION}$ weather file produced with location-based weighting factors, and a TMY_{CZ} weather file produced with climate zone-based weighting factors. The location-based weighting factors are produced using the methodology outlined in Chapter 2. The climate zone-based weighting factors are determined by taking the average of the location-based weighting factors generated for each city within the climate zone.

Six prototypical medium office building energy models are developed based on the NECB 2020 requirements for each climate zone and are simulated with the hourly historical weather data for each location. An ideal air loads model is used to assess the building energy demand. The $TMY_{STATION}$ and TMY_{CZ} weather files are compared to the CWEC weather files using the long-term average weather data and two performance indicators, the coefficient of variation of the root mean square error (CV(RMSE)) and normalized mean bias error (NMBE) to evaluate the results. The CV(RMSE) evaluates the variation in the model and the NMBE evaluates the bias in the model. Furthermore, the ASHRAE Guideline 14 provides acceptable ranges for both the CV(RMSE) and NMBE to determine if the results are acceptable.

The $TMY_{STATION}$ and TMY_{CZ} weather files performed better than the CWEC weather files in representing the long-term average for the majority of the locations as indicated by lower CV(RMSE) and NMBE values. In the few cases where CWEC had a lower NMBE value compared to the $TMY_{STATION}$ and TMY_{CZ} weather files, the CV(RMSE) value was lower for the proposed TMY weather files indicating the outliers in the CWEC results may have cancelled out to provide a better NMBE value. Additionally, CWEC had a total of three locations with a lower CV(RMSE) however the performance was at most 0.3% better while the generated TMY weather files were up to 5.1% better.

The $TMY_{STATION}$ weather files had the most instances where both the CV(RMSE) values and NMBE values were the smallest. The improved representation of the long-term average weather data across the varying climates demonstrates that the methodology proposed in manuscript #1 is versatile and can be applied to various climates. The TMY_{CZ} weather files present a viable alternative, with the NMBE and CV(RMSE) values being very similar to the $TMY_{STATION}$ weather files for most locations, with up to a 0.9% difference in performance between the two proposed weather files. The time saved by using the climate zone-based weighting factors to generate TMY weather files may be worth the marginal trade-off in performance between the $TMY_{STATION}$ and the TMY_{CZ} weather files. Although the TMY_{CZ} weather files demonstrated an improved representation in the long-term average compared to the CWEC weather files, for locations with weather patterns which significantly differ from other cities within the same climate zone it is recommended to use the $TMY_{STATION}$ files for these locations until the climate zone definitions are improved.

Although both proposed TMY weather files performed better than the CWEC weather files, the study had a few limitations. The study focused on one building type and was limited to Canadian climate zones. Furthermore, the standardized climate zone-based weighting factors were calculated based on a small sample for each climate zone. The limitations present opportunities for future work which may improve the methodology proposed in Manuscript #1 for cases of varying applications and climates. Lastly the definitions of Canadian climate zones should be improved to account for other variables such as cooling degree days. While climate zone-based weighting factors enhance the potential adoption of the proposed methodology, further research is crucial to redefine the climate zone definitions and expand the study to encompass diverse building types, sizes, and a broader spectrum of cities.

Chapter 2 Defining generation parameters with an adaptable data-driven approach to construct typical meteorological year weather files.

2.1 Contribution of Authors

This manuscript is published in the Journal of Energy and Buildings. The journal paper proposes a machine learning methodology to improve the current process of TMY weather file generation by using a data-driven approach to determine the weather parameters and relevant weighting factors used in the Sandia method. Ashleigh Papakyriakou is the second author of the paper. Ashleigh's contributions include the conceptualization, development and validation of the energy model, writing, editing and review. Anahita Bigtashi (first author) contributions include the conceptualization, development of the methodology and proposed weather files, data curation, formal analysis, writing, editing and review. Dr. Bruno Lee contributed to conceptualization, and review.

A. Bigtashi, A. Papakyriakou, and B. Lee, "Defining generation parameters with an adaptable data-driven approach to construct typical meteorological year weather files," *Energy and Buildings*, vol. 303, p. 113781, 2024. doi:10.1016/j.enbuild.2023.113781

2.2 Introduction

A significant emphasis has been placed on energy efficiency and reliability due, in part, to the shift in the climatic condition brought on by anthropogenic climate change. This increase in demand has drove the building industry to shift towards creating high-performance buildings which focus on reducing building energy consumption. In order to adequately design high-performance buildings, a thorough evaluation of the building envelope, and the mechanical, electrical and renewable energy systems, is required. As a result, energy simulation tools, often referred to as energy models, have become widely adopted by designers to estimate the energy performance of renewable energy systems and buildings. Weather files are a critical component in energy simulation as they provide essential information on the environmental conditions to which the building or system is exposed.

2.2.1 Typical Meteorological Year Weather Files

Energy simulations are typically performed using reference year (RY) weather files to reduce the computational time associated with multi-year simulations. These weather files are created from long-term weather data which is synthesized into a single year using different statistical methods. The use of RY weather files allows designers to evaluate different design configurations in a timely manner. However, the accuracy of the simulation results relies heavily on the selected RY weather file and its ability to adequately reflect the intricacies of the long-term weather data.

RY weather files may be generated to reflect extreme weather conditions, known as extreme year weather files, or to represent the average weather conditions, known as typical meteorological year (TMY) weather files. Extreme weather files are often used to determine how a design will perform under extreme weather conditions. TMY weather files are meant to reflect the average condition for the selected period, disregarding extreme weather conditions, with the intent to show the long-term average performance [1]. RY weather files may be generated using historical weather data or, future weather data obtained from existing climate change models.

In building energy simulation, typical meteorological year weather files are generated using, historically, a 30-year period, which is synthesized into a single year. There are a few common methods used to generate TMY weather files, such as the Danish method [2], the Festa and Ratto method [3], the Sandia National Laboratories method [4] [5] and the ISO 15927-4 standard [6].

The Sandia method is one of the most frequently used methods to generate TMY weather files. Many studies conducted at the start of the industry's push in building energy simulation have found the Sandia method to be an adequate representation of the long-term average weather data [7] [8] [9] [10]. The Sandia method is used to generate many commonly used RY weather files, including; Typical Meteorological Year ([TMY][1], TMY2, TMY3) [4] [11] [12], International Weather Year for Energy Calculations (IWEC, IWEC2) [13] [14] and Canadian Weather Year for Energy Calculation (CWEC) [1] [15].

[1] *[TMY]:Used to denote the Typical Meteorological Year file format originally presented in* [2].

The Sandia method is an empirical approach that selects individual months from different years within the long-term weather dataset to create the TMY weather file [12]. Each month within the long-term weather dataset is evaluated on a pre-defined set of weather parameters, referred to as decision weather parameters. The decision weather parameters are attributed a weight based on their perceived importance. The months found to be the most statistically similar to the long-term average weather dataset are retained and concatenated into a single year [12]. The decision weather parameters and their corresponding weights vary between different TMY weather file formats. The detailed outline of the Sandia method is presented below, with specific terminology outlined in Table 2.1.

For each month of a year:

i. The average/total daily long-term and short-term (candidate month) cumulative distribution function (CDF) is obtained for each of the decision weather parameters outlined in Table 2.2.

ii. For each candidate month within the long-term dataset, the corresponding short-term CDF is compared to the long-term CDF using the Finkelstein-Schafer (FS) statistic, as outlined in equation (2.1). In this equation, the sum of the absolute difference between the short-term (candidate month) and long-term CDF is obtained. The process is repeated for each decision weather parameter.

$$FS_j = \left(\frac{1}{n}\right)\sum_{i=1}^{n}\delta_i \tag{2.1}$$

FS_j: *FS value for decision weather parameter j*

δ_i: *Absolute difference between the short-term (candidate month) and long-term CDF*

i: *Day of the month*

j: *Decision weather parameter*

n: *Total days in the month*

iii. The weighted sum (WS) is calculated for each candidate month according to equation (2.2). The weighting factor assigned to each decision weather parameter, as outlined in Table 2.2 is multiplied by the FS value obtained in step ii.

$$WS = \sum_{j=1}^{n} w_j FS_j \tag{2.2}$$

j: Decision weather parameter

n: Total number of weather parameters

FS_j: FS value for decision weather parameter j

w_j: Weighting factor for weather parameter j

iv. The candidate months are ranked in ascending order of their WS and the top five candidate months are retained.

v. The five candidate months are re-ranked according to the proximity to the long-term mean and median values for dry-bulb temperature and global horizontal irradiance.

vi. The persistence of mean dry-bulb temperature and daily global horizontal radiation are evaluated by determining the frequency and length of consecutive days with measurements outside the fixed long-term percentiles. Each candidate month is evaluated based on the number of consecutive days outside of the following three cases:

a) below the 33rd for dry-bulb temperature

b) above the 67th percentile for dry-bulb temperature and

c) below the 33rd percentile for global horizontal irradiance.

The first candidate month that meets the persistence criteria is selected as the final candidate month.

Each of the outlined steps is repeated for each month in a calendar year until all 12 selected candidate months are concatenated into a single year.

Table 2.1: Sandia method terminology for TMY weather file generation

Terminology	**Description**
Candidate month	The month of each individual year within the long-term dataset currently being evaluated. (ex: January 2001, January 2002, January 2003)
Decision weather parameter	The weather parameters considered in the FS statistical method and WS to generate the TMY weather file.
Long-term CDF	The cumulative distribution function for the entire period of the weather dataset for the calendar month currently being evaluated. (ex: all January's within the 20 year dataset)
Generation parameters	The weather parameters and their corresponding weights considered in the FS statistical method and the WS to generate the TMY weather file, as outlined in Table 2.
Short-term CDF	The cumulative distribution function for the candidate month currently being evaluated. (ex: January 2001)

Table 2.2 presents the decision weather parameters used for different TMY weather file formats and their corresponding weighting factors. The weights presented in Table 2.2 have been normalized to facilitate comparison.

Table 2.2: Typical meteorological year weather file weighting factors

Weather Parameter	*Measurement Parameter*	*CWEC [1][15] IWEC [13]*	*[TMY][1][4]*	*TMY2 [11], TMY3 [12], IWEC2 [14]*
Dry-bulb temperature	Mean daily	0.300	0.083	0.100
	Minimum daily	0.050	0.042	0.050
	Maximum daily	0.050	0.042	0.050
Dew point temperature	Mean daily	0.050	0.083	0.100
	Minimum daily	0.025	0.042	0.050
	Maximum daily	0.025	0.042	0.050
Wind Speed	Mean daily	0.050	0.083	0.050
	Maximum daily	0.050	0.083	0.050
Global horizontal irradiance	Total daily	0.400	0.500	0.250
Direct normal irradiance	Total daily	-	-	0.250

The weighting factors used for CWEC, IWEC, IWEC2, TMY2 and TMY3, have been modified from the original [TMY][1]. The original [TMY][1] weighting factors were selected for a solar heating system and assigned according to experts' judgement [4]. In their study, Hall et al. [4] further stipulated that the decision weather parameters and their corresponding weights are application specific. Therefore, different weather parameters and/or weights may be required for different applications, such as building energy simulation. TMY2 modified the weighting factors for dry-bulb temperature, dew point temperature, and wind speed to place a greater emphasis on both dry-bulb and dew point temperature [11]. Direct normal irradiance was added as it was found to improve the comparison between the TMY2 and the 30-year annual average [11]. The TMY2 user's manual indicates that these weather files are intended to be used for simulations of solar energy conversion systems and building systems. The manual further specifies that these files may not be appropriate for simulations of wind energy conversion systems. The CWEC weather file weighting factors were assigned based on the assumed influence that various weather parameters have on building energy usage [1][15].

2.2.2 TMY Generation Limitations and Constraints

Numerous studies have investigated the impact of generation parameters on simulation results for different applications and climate. Current TMY generation methods rely on universal weighting factors defined based on expert judgement, which often neglects variations in climate and application. Furthermore, for most generation methods, the generation parameters used to produce TMY weather files are often pre-defined, which poses a significant constraint for regions with limited weather data.

2.2.2.1 Varying applications

As outlined by Hall et al. [4] in their original publication, the importance attributed to decision weather parameters may vary based on the purpose of the application and the type of investigation. Chan [16] used a genetic algorithm (GA) to develop weighting factors to generate TMY weather files using the Sandia method. The decision weather parameters selected for the study were those used to generate IWEC weather files. The study evaluated four different applications: a fully air-conditioned office building, a building attached non-concentrating photovoltaic (BaPV) system, a wind turbine power generation system, and a concentrating solar power (CSP) system. Chan [28] [16] determined the set of weighting factors obtained for the fully air-conditioned building to be closest to those used in the IWEC file generation. In contrast, the weights obtained for the three other applications significantly differed from the IWEC weighting factors. Furthermore, the simulation results for the application specific TMY weather files were determined to better reflect the long-term mean than the IWEC simulation results [16]. The results of the study further highlight the potential need for weighting factors tailored for different applications.

Additionally, Georgiou et al. [17] completed a sensitivity analysis of the impact of weighting factors on TMY weather file generation for various applications. The investigation was implemented for a residential solar thermal system, a wind turbine generator, and the heating and cooling analysis of a typical dwelling in Cyprus. The study found that TMY weather file datasets can significantly vary due to the assigned weighting factors. The authors concluded that weighting factors should be optimized for the intended use.

Kambezidis et al. [18] conducted a study in which TMY weather files were generated for different applications using the modified Sandia National Laboratories method for 33 different locations in

Greece. The TMY weather files were generated for five different applications: meteorology-climatology, biometeorology, agro-meteorology-hydrology, PV applications, and building energy. The different weighting factors used to generate the TMY weather files for each application were selected based on experts' judgment and previous studies. The results showed good agreement between the application specific TMY weather files and the long-term average simulation results, further demonstrating the impact of application-based weighting factors on simulation accuracy.

Yang, H and Lu, L [19] investigated the impact of different typical meteorological year (TMY) and example weather year (EWY) weather files on simulation results for building energy and renewable energy systems. The investigation was performed for a commercial building and a hybrid solar – wind power system in Hong Kong. The TMY and EWY weather files were generated for Hong Kong using variations of the Sandia method. The decision weather parameters and weights used to generate the weather files were selected based on existing publications. Yang, H and Lu, L [19] concluded that different applications require different weighting factors as each case exhibited varying degrees of deviations, with the solar – wind system being the most significant.

Finally, in their study, Sun et al. [20] highlighted the lack of TMY weather files developed for daylight-utilized building energy simulation. As current TMY weather files are not adequate for daylighting analysis, the authors use a genetic algorithm to create optimized weighting factors. The resulting TMY weather files are generated using the Sandia method and the optimized weighting factors. In addition to the existing weather parameters used in the Sandia method, the daily mean and maximum global solar irradiance, direct normal irradiance and horizontal diffuse irradiance were added.

2.2.2.2 Varying climates

As previously established, most commonly available TMY weather files are generated using decision weather parameters with weights assigned based on expert judgment. Despite differences in the climatic condition for different regions around the world, standard TMY weather files adopt universal weighting factors. These universal weighting factors not only disregard the differences in climate between different regions, but also ignore seasonal variations present in weather data.

A study by Meng, F et al. [21] investigated the impact of different weather parameters on the heating energy consumption of an office building for different climate zones across China. The study determined dry-bulb temperature as having the greatest influence on the building heating energy consumption for all considered climate zones. Furthermore, although other weather parameters were noted to have a slight influence on the consumption, the significance was found to vary depending on the climate zone [21]. Kalamees et al. [22] conducted a similar study investigating the impact of different weather parameters on building energy demand for different cold boreal climates. Sensitivity tests were used to determine the influence of the different weather parameters for various seasons. The study determined dry-bulb temperature as having the greatest influence for all climates. Furthermore, in summer, solar irradiance was found to have a significant impact of building energy consumption. Both studies highlight the differences in significance attributed to certain weather parameters for varying climates and seasons.

Hong et al. [23] employed large-scale building simulation to investigate the impact of climate data on peak electricity demand and energy use for 17 ASHRAE climate zones. In this study, the simulation results obtained using 30 years of historical weather data are compared to those obtained using TMY3 for three types of office buildings. The study found significant discrepancies between the TMY weather file and long-term weather data simulation results. The impact was found to vary based on the office building size, energy efficiency level, and climate; with significant discrepancies noted for cold climates. In a similar study, Seo et al. [10] completed a sensitivity analysis on the impact of different TMY weather file selection procedures on building energy analysis for varying U.S. climates. The study assessed different TMY weather file generation procedures, weighting factors and historical weather data periods using building energy simulation. The difference in building energy use and peak demand between the long-term average historical weather data and the investigated TMY weather files simulation results were compared. The study found the largest differences to be for heating energy use and warm climates. In the case of both studies, the results further demonstrated the variability in performance of TMY weather files for different climates. Therefore, to improve the accuracy of TMY weather files for different climate zones, optimized weighting factors are required.

As a means to address the issue of universal weighting factors adopted by most readily available TMY weather files, Li et al. [24] proposed a new method to generate TMY weather files using the entropy-based Technique for Order Preference by Similarity to Ideal Solution (TOPSIS) theory. In this study, the authors generated TMY weather files for five different climate zones in China using the same decision weather parameters outlined in the Sandia method. In the method outlined by Li et al. [24], both the weights and resulting weather file are generated using the objective weather data and, therefore, account for climate-dependent variations. The results obtained in the study found that the TMY weather files generated using their proposed method more closely reflected the long-term average historical weather data. This study demonstrates the potential improvement in accuracy associated with climate-based weighting factors and further highlights the issue associated with judgement-based generation parameters.

2.2.2.3 TMY Generation Constraints

Most TMY weather files generated using the Sandia method use universal generation parameters defined based on expert judgement, which often disregard variations in applications and climate. Furthermore, the use of generalized decision weather parameters and weighting factors poses a significant constraint for areas with limited weather data. While satellite data has become widely available, access to long-term historical weather data with adequate temporal resolution for all expert-defined weather parameters may be limited in some regions.

Many studies have proposed different modifications of the Sandia method to address some of the presented issues, as well as improve the resulting TMY weather file. A preliminary investigation presented by Gai et al. [25] proposed a simplified method, in which, only the mean daily value for each decision weather parameter is considered. The method was proposed to address the issue of inadequate temporal resolution of historical weather data for certain rural regions in China. Ohunakin et al. [26] sought to generate TMY weather files tailored to the climate of north-east Nigeria using the Sandia method. The study relied on expert judgement to define a new set of weighting factors considered to adequately reflect the designated climate and application. Finally, Zang et al. [27] proposed a hybrid method using a combination of the Danish method, Festo-Ratto method, and a modified Sandia method to generate improved TMY weather files. In this approach, the candidate months obtained for all three generation methods are evaluated based on their proximity to the long-term average data. The candidate months found to have the least difference to the long-term average are then selected to create the TMY weather file.

2.2.3 Machine Learning Applications

The uncertainty associated with the use of judgement-based generation parameters is partly due to the absence of a data-driven approach for defining decision weather parameters and weighting factors. Therefore, it is crucial to develop a data-driven approach that systematically identifies weather parameters and their respective importance which considers climate variations and application-specific considerations.

As previously mentioned, a limitation of current TMY generation methods is their reliance pre-defined decision weather parameters. In the case of varying applications, weather parameters considered to have a significant impact on simulation results may be disregarded in the TMY generation process. Therefore, decision weather parameters should be selected based on their relevance to the investigated application.

In machine learning (ML), variables deemed to have little significance on the target variable are typically disregarded in order to enhance the performance of the model. In supervised learning, a machine learning algorithm analyzes the relationship between predictor variables and the target variable to construct a predictive model. In this process, known as training, datasets are fed to the ML algorithm to build the model. The training datasets are typically divided into two parts: the training input, consisting of independent variables known as features, and the training output, comprising the target variable. A common misconception in machine learning is that larger input datasets with numerous features will result in a better model. However, this approach can often introduce excess non-contributive features, also known as noise, which can degrade the model's performance. Feature selection is the process of filtering the training input dataset to retain only relevant features, thereby reducing noise and improving the quality of the data. Certain tree ensembles, such as bagging, random forest and boosting, have been found effective in retaining relevant features, however, these types of algorithms are also vulnerable to noise [28].

Another limitation of current TMY generation methods is the use of judgement-based universal weighting factors, which fail to consider variations in climate and applications. By contrast, the use of machine learning (ML) provides a data-driven approach to determine the importance of variables in predicting outcomes, known as feature importance. ML algorithms such as gradient boosted tree algorithms and Random Forest are particularly effective in identifying predictorresponse relationships and assessing the importance of each feature [29] [28][29]. Thus, ML algorithms may be used to develop a data-driven approach to obtain weighting factors which account for changes in climate and application. For instance, Hosseini et al. [30] devised a ML-based approach to define weighting factors using tree ensembles. In the proposed approach, a Random Forest regression model is used to extract the feature importance for the nine decision weather parameters used to generate CWEC files. The study investigated five different training approaches using building energy demand to determine the most suitable training output parameter and weighting factor resolution. The findings revealed that the TMY weather files generated using the extracted weights better reflected the long-term average building energy demand compared to CWEC for regions with significant temperature fluctuations [30]. Additionally, although the five approaches yielded similar results, the TMY weather files generated using monthly weighting factors demonstrated slightly better performance. While the presented approach offers a systematic way of determining weighting factors, the methodology remains difficult to adapt for varying applications due to the use of pre-defined weather parameters. Furthermore, the issue of limited weather data, both in terms of available weather parameters and resolution, remains unaddressed.

Current TMY generation methods rely on judgement-based universal generation parameters. Machine learning provides a data-driven approach to determine both the relevant weather parameters and their corresponding weights. The purpose of this study is to improve the current approach to TMY weather file generation by providing a data-driven framework to define the generation parameters which account for climate and application. Building upon the study by Hosseini et al. [30], the proposed methodology utilizes a ML regression algorithm to systematically identify the decision weather parameters and weighting factors used in the TMY generation process. The weather parameters and weighting factors are subsequently integrated into the Sandia method to generate the TMY weather files. The data-driven approach is applied in a case study investigating building energy demand of a prototypical office building in Montreal, Canada.

The methodology aims to overcome the uncertainty associated with judgement-based generation parameters for diverse climates and applications, as well as to address existing constraints related to weather data availability. In order to improve weather file adaptability, the proposed methodology leverages machine learning and simulation results to define the generation parameters. Consequently, while the present case study centers on building energy demand for a specific locale, the data-driven approach may be applied for a variety of climates and applications by utilizing corresponding simulation results to train the ML regression model.

2.3 Methodology

The presented methodology uses a data-driven ML approach to define the generation parameters used to produce TMY weather files. The proposed approach is divided into four stages as outlined in Table 2.3. Each stage is presented in a flowchart at the beginning of each corresponding section. A flowchart outlining the entire combined methodology is provided in Appendix A.

In Stage 1, long-term weather data and building energy demand results are acquired and processed. In Stage 2, the decision weather parameters are obtained using feature selection. In this stage, the datasets obtained in Stage 1 are evaluated using a gradient boosted tree regression model to identify the weather parameters that have the greatest influence on building energy demand. In Stage 3, the feature importance of each weather parameter from Stage 2 is extracted and used as weighting factors, in Stage 4, to generate a TMY weather file.

Table 2.3: Outline of proposed methodology

Stage			
1	Training dataset acquisition and processing		
		Step	
		1.1	Acquire long-term weather data
		1.2	Simulate long-term building energy demand
		1.3	Filter weather data and remove multicollinearity
2	Generation parameters: decision weather parameters		
		2.1	Train the gradient boosted tree regression model
		2.2	Select *decision weather parameters*
3	Generation parameters: monthly weighting factor		
		3.1	Retrain the gradient boosted tree regression model
		3.2	Select *monthly weighting factors*
4	Typical meteorological year weather file generation		
		4.1	Generate the TMY weather file using the Sandia method

Stage 2 and 3 of the proposed methodology constitute the primary contribution of this study by presenting a systematic data-driven approach to selecting decision weather parameters and weighting factors. The purpose of this section is to present the methodology in generic terms to facilitate reproducibility for varying climate and applications. Additional details are provided on the exact procedures implemented in this study in section 2.4 by means of a case study.

2.3.1 Stage 1: Training Dataset Acquisition and Processing

The purpose of Stage 1 is to acquire and process the datasets used to train the machine learning regression model presented in Stage 2 and 3. Figure 2.1 presents an overview of each step in Stage 1, including the resulting outputs used in Stage 2 and 3.

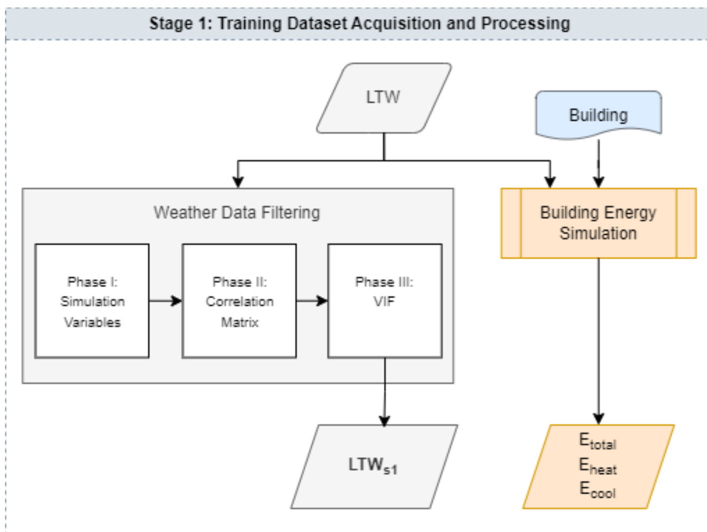

Figure 2.1: Stage 1 from the proposed methodology flowchart

2.3.1.1 Step 1.1: Weather data

The proposed methodology uses long-term weather (LTW) data to perform multi-year energy simulations, train the machine learning regression model, and generate the TMY weather files. Therefore, hourly long-term weather data for the location of study is required. The suggested minimum period is 15 years, however, in cases of limited data, shorter periods may be used. Furthermore, the dataset must include all weather parameters required by the selected simulation software for the application of choice. In this case, long-term hourly weather data, which includes all weather parameters required for building energy simulation, is acquired for a period of 20 years. Additional information is provided on the long-term weather dataset in section 2.4.1.

2.3.1.2 Step 1.2: Building energy simulation

Building energy simulation is performed using the long-term weather (LTW) data acquired in step 1 and the resulting hourly heating and cooling energy demand is extracted. The hourly energy demand results are used as output datasets to train the regression model, as presented in section 2.3.2 and 2.3.3.

2.3.1.3 Step 1.3: Data filtering and multicollinearity

As previously mentioned, certain machine learning algorithms, are vulnerable to noise [28]. In order to address this issue, as well as model-specific issues, an initial filtering process is conducted to remove excess features (weather parameters) from the long-term weather data. The filtering process is performed in three phases. In each phase, the previously retained features (weather parameters) are evaluated and removed based on the below-outlined criteria.

Phase I: Simulation Variables

Long-term weather datasets often include additional weather parameters which are, depending on the application, unnecessary for simulation. Moreover, the inclusion of categorical and discrete variables can significantly impact the performance of ML regression models. Therefore, in Phase I, weather parameters categorized as discrete variables or deemed unnecessary for simulation are considered as excess features and removed. In other words, to minimize noise within the training input dataset, the long-term weather data is initially filtered to only retain weather parameters with continuous data which are considered by the selected simulation software.

Phase II: Correlation Matrix

As briefly mentioned, the present study seeks to use a gradient boosted tree regression model to determine the feature importance scores for each predictor variable (weather parameter). Although an initial filtering is performed to remove noisy data from the training input dataset, the issue of multicollinearity remains. In the case of tree-based models, the presence of highly correlated predictors within a dataset creates a redundancy which dilutes their importance scores [28]. To address this issue, the Phase I dataset is filtered, once again, using a combination of Pearson's correlation coefficient (Phase II) and variance inflation factor (Phase III). In Phase II, weather parameters with an absolute correlation of 0.75 or above are deemed to be highly correlated and are removed.

Phase III: Variance Inflation Factor (VIF)

In Phase III, the remaining weather parameters are evaluated using the variance inflation factor (VIF). While Phase II addressed the issue of collinearity, the variance inflation factor is used to detect multicollinearity, as it considers the relationship between a single variable to a group of variables. An initial evaluation is made using all remaining weather parameters. A VIF score exceeding 10 is considered to indicate high correlation. The weather parameter with the highest VIF score above the threshold of 10 is removed. The process is repeated until all remaining parameters fall below the defined threshold.

Following Phase III, a subset of the long-term weather data is created using the remaining weather parameters. The resulting long-term weather dataset subset (LTW_{S1}) is used in Stage 2 to train the regression model. Additional details on the filtering results obtained in this study are presented in step 3 of section 2.3.2.1.

2.3.2 Stage 2: Generation Parameters: Decision Weather Parameters

The purpose of Stage 2 is to identify the weather parameters which have a significant impact on building energy demand using feature selection. The identified weather parameters are used as the decision weather parameters to generate the TMY weather files outlined in section 2.2.1. The retained decision weather parameters are used in Stage 3 to retrain the regression model. An overview of the procedures and datasets considered in this stage is presented in Figure 2.2.

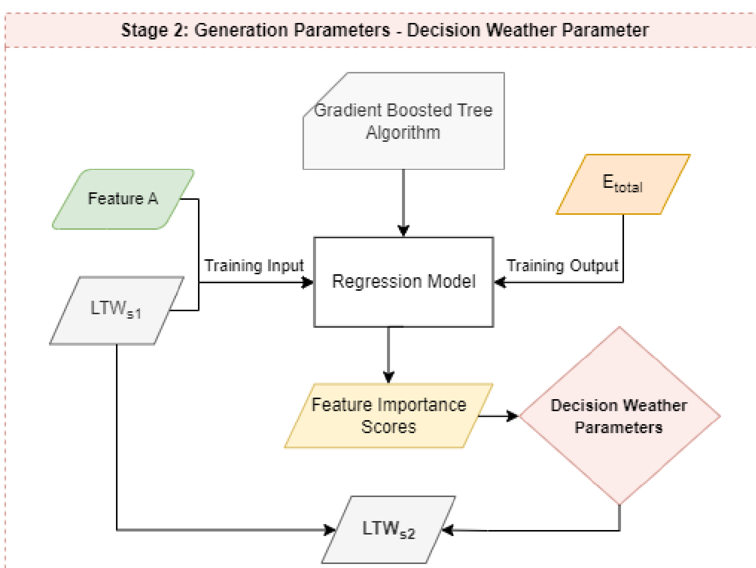

Figure 2.2: Stage 2 from the proposed methodology flowchart

2.3.2.1 Step 2.1: Decision tree regression model training

In Stage 2, a gradient boosted tree regression model is trained using long-term hourly weather data and, for this study, building energy demand. In the case of varying applications, alternative simulation results, considered continuous in nature and suited to the application of study, may be used. The model is used to extract the feature importance for each of the weather parameters obtained in Stage 1 (LTW_{S1}). The purpose of the feature importance scores, at this stage, is to evaluate and help identify the relevant or "important" weather parameters within LTW_{S1}. In other words, the feature importance is obtained for each weather parameter and used for feature selection.

Given the absence of a universal cut-off value for feature importance scores, a threshold must be established to identify the relevant features within the dataset. As such, an additional feature, comprised of a set of randomly generated numbers, is incorporated into the training input dataset. The purpose of the additional feature, referred to as Feature A, is to set a minimum threshold for the feature importance scores.

As a result, the gradient boosted tree regression model is trained using LTW_{S1} and Feature A as training input datasets and total hourly energy demand (E_{total}) as the training output dataset. Additional details on Feature A are provided in Appendix A.

2.3.2.2 Step 2.2: Decision weather parameters

Once the model is trained, the monthly feature importance score is extracted for Feature A and each weather parameter within LTW_{S1}. In other words, a separate feature importance score is obtained for each month.

As previously established, the purpose of Feature A is to establish a minimum threshold or "cut-off" to identify the relevant features. Thus, the monthly feature importance scores obtained for each weather parameter are compared to those obtained for Feature A. The weather parameters with feature importance scores exceeding those of Feature A are retained to create a new long-term weather data subset (LTW_{S2}).

2.3.3 Stage 3: Generation Parameters: Monthly Weighting Factor

In Stage 3, the model is retrained using only the retained features (decision weather parameters) from Stage 2 (LTW_{S2}) to obtain undiluted features importance scores, as outlined in Figure 2.3. The purpose of this stage is to determine the appropriate weights to attribute each decision weather parameters using their corresponding feature importance score.

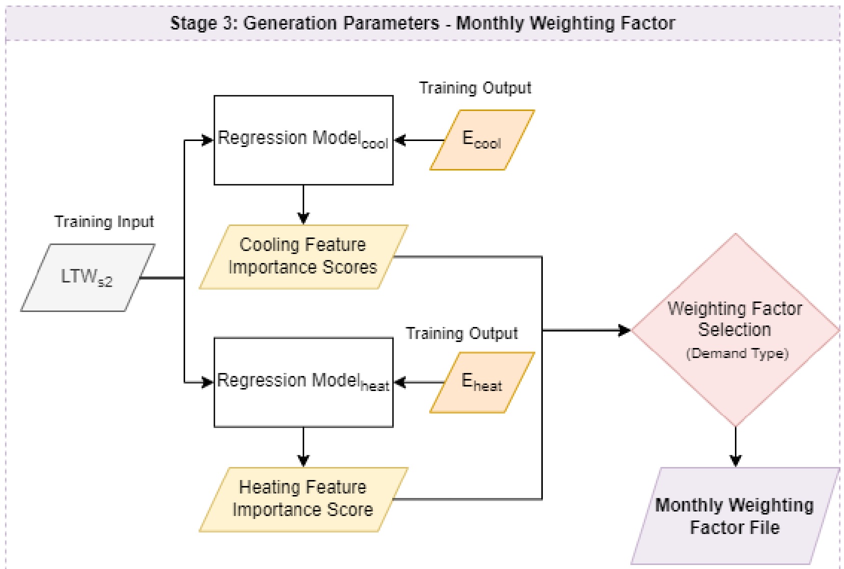

Figure 2.3: Stage 3 from the proposed methodology flowchart

2.3.3.1 Step 3.1: Decision tree regression model retraining

In this step, the gradient boosted tree model is retrained using the long-term weather data subset LTW_{S2}. As previously done in Stage 2, the hourly building energy demand is used as the training output dataset. However, in Stage 3, the training output dataset is separated into heating (E_{heat}) and cooling (E_{cool}) demand.

2.3.3.2 Step 3.2: Monthly weighting factors

As previously mentioned, in Stage 3, the model is trained using heating and cooling energy demand, resulting in two models. In both cases, a monthly importance score is obtained from each model using the built-in feature importance function. This results in two sets of 12 feature scores, one for heating and one for cooling, for each feature.

In order to generate a single set of monthly weighting factors for each of the decision weather parameters, a subset of the 24 feature scores must be selected. In this case, the average long-term heating and cooling demand is obtained for each month. The months are then classified as either heating or cooling dominant. The feature scores are then selected for each month based on the month's heating or cooling demand classification, and output into a single file. In other words, based on the classification of the month, the corresponding (heating or cooling) feature score is selected. The resulting file is composed of monthly weighting factors (feature scores) for each of the decision weather parameters. The resulting monthly weighting factor file is used in Stage 4 as the attributed weights for the decision weather parameters to calculate the weighted sum.

2.3.4 Stage 4: Typical Meteorological Year Weather File Generation

In this stage, the decision weather parameters defined in Stage 2, and their corresponding monthly weighting factors determined in Stage 3, are integrated into the Sandia method to generate a typical meteorological year weather file (Figure 2.4).

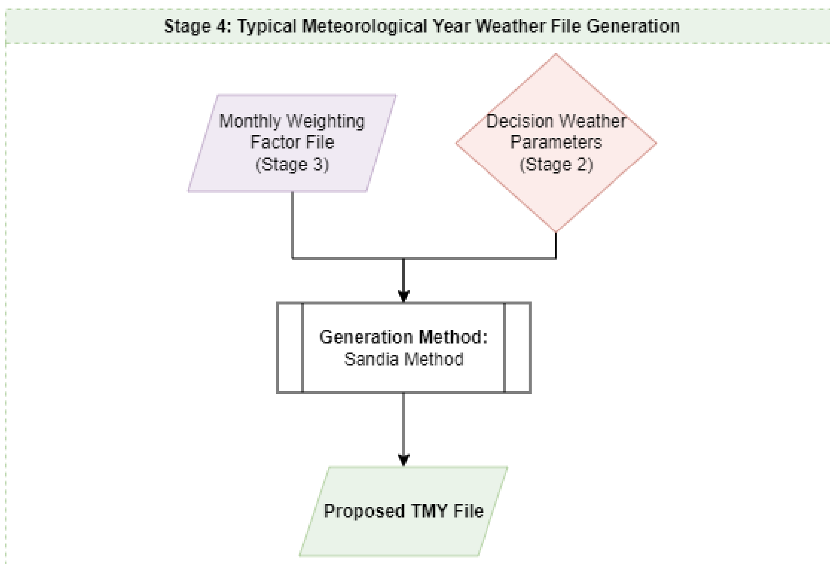

Figure 2.4: Stage 4 from the proposed methodology flowchart

2.3.4.1 Step 4.1: Sandia method integration

The typical meteorological year weather file is generated using the original long-term weather data (LTW). The long-term weather data is synthesized into a single-year weather file using the steps outlined in the Sandia method, as presented in section 2.2.1.

The decision weather parameters and their corresponding monthly weighting factors, outlined in Stage 2 and 3, are used in steps i to iii of the Sandia method. In the Sandia method, each month is evaluated individually. Therefore, although the initial Sandia method publication utilizes only a single set of weighting factors, no modification to the procedure is needed to accommodate the use of monthly weighting factors.

2.4 Case Study

The case study used to present the proposed methodology is for a medium office building located in Montreal, Canada. In this section, details are provided regarding the specific datasets, statistical tests and software used for each step of the proposed methodology, as well as an overview of the performance indicator used to evaluate the generated TMY weather files.

2.4.1 Stage 1: Montreal Medium Office Simulation and CWEEDS Filtering

In this study, the historical long-term weather (LTW) data for the Montréal-Pierre Elliott Trudeau International Airport is obtained from the Canadian Weather Energy and Engineering Datasets (CWEEDS). The CWEEDS dataset is composed of hourly data for a period of 20 years, spanning between 1998 to 2017. The CWEEDS data was selected to facilitate the comparison between CWEC and the proposed TMY weather file generated in this study. Furthermore, the use of CWEEDS allows to avoid potential discrepancies stemming from using different long-term weather data which may result in differing periods, resolutions, and measurements.

The building energy simulation of a prototypical medium office building is performed in EnergyPlus v.23.1.0 using the unfiltered CWEEDS data. The building geometry was modelled after the DOE medium office building prototype [31] and modified according to the 2020 National Energy Code of Canada for Buildings (NECB) [32] requirements to adhere to Canadian building standards. The building model parameters are presented in Table 2.4. The heating and cooling setpoints were set to 21°C and 24°C, respectively. Finally, an ideal air loads system was modelled to represent an ideal HVAC system. The building energy demand results obtained using the long-term weather (LTW) data is used to train the regression model, as well as provide a baseline to compare the performance between CWEC and the proposed TMY weather file.

Table 2.4: Medium office building model parameters [32].

Building Parameter	***Value***	***Units***
Wall U-value	0.240	W/m^2K
Roof U-value	0.138	W/m^2K
Slab on Grade U-value	0.757 for 1.2m	W/m^2K
Window U-value	1.73	W/m^2K
WWR	0.39	-
Infiltration	0.25	$L/s/m^2$ at 5 Pa
Lighting	10.00	W/m^2
Receptacle Equipment	7.50	W/m^2
Occupancy	25.00	m^2/person
Schedule	Schedule A	-
Conditioned Floor Area	6,898.00	m^2

In this stage, an initial filtering of the training input dataset (CWEEDS) is performed in three phases, as outlined in section 2.3.1. The results of the filtering process are presented in Table 2.5.

Table 2.5: Long-term weather data parameter and filtering overview

Weather Parameter	***Abbreviation***	***Phase I: EnergyPlus Variables***	***Phase II: Correlation Matrix***	***Phase III: VIF***
Dry-bulb temperature	DBT	✓	✓	✓
Dew point temperature	DWPT	✓	**X**	**X**
Relative humidity	RH	✓	✓	✓
Atmospheric Pressure	ATM	✓	✓	**X**
Direct normal irradiance	DNI	✓	✓	✓
Diffuse horizontal irradiance	DHI	✓	✓	✓
Wind Speed	WS	✓	✓	✓

In Phase I, the LTW data extracted from CWEEDS is first filtered to only retain continuous weather parameters considered by EnergyPlus. From this phase, the following seven weather parameters are retained and further evaluated in Phase II: dry-bulb temperature, dew point temperature, relative humidity, atmospheric pressure, direct normal irradiance, diffuse horizontal irradiance and wind speed.

In Phase II, a correlation matrix (Figure 2.5) is constructed using Pearson's correlation coefficient for the weather parameters retained in Phase I. The correlation coefficient obtained for dry-bulb temperature and dew point temperature was determined to be 0.94, which exceeds the previously established threshold of 0.75. As a result, the dew point temperature is removed from the dataset prior to Phase III.

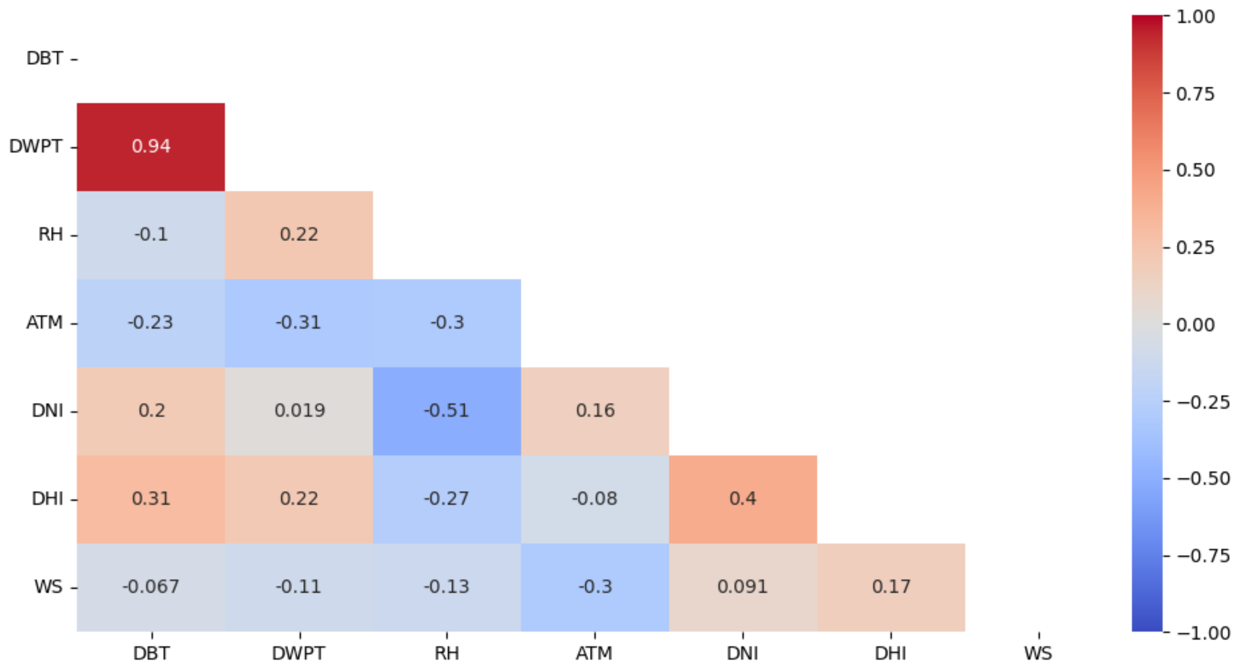

Figure 2.5: Pearson correlation matrix of long-term weather data

In Phase III, the variance inflation factor (VIF) scores are obtained for the remaining six weather parameters and presented in Table 2.6. In this case, atmospheric pressure is found to have the greatest VIF score and, as a result, is removed from the dataset. The VIF scores are re-calculated for the remaining weather parameters, the results of which are presented in Table 2.5. The resulting VIF scores obtained for the filtered predictors fall below the pre-set threshold of 10.

Table 2.6: VIF Scores for Phase III weather parameters

Weather Parameter	***Initial VIF Scores***	***Filtered VIF Scores***
Dry-bulb temperature	1.60	1.58
Relative humidity	23.95	3.45
Atmospheric Pressure	31.96	-
Direct normal irradiance	2.03	1.60
Diffuse horizontal irradiance	1.99	1.97
Wind Speed	4.16	3.63

Following Phase III, a subset of the CWEEDS data is created to solely include the five retained weather parameters: dry-bulb temperature, relative humidity, direct normal irradiance, diffuse horizontal irradiance and wind speed. The newly created CWEEDS subset is referred to as LTW_{s1} and used as the training input dataset for the machine learning regression model in Stage 2.

2.4.2 Stage 2: Decision Weather Parameters Selection Using XGBoost

As outlined in step 4 of section 2.3.2 of the methodology, a gradient boosted tree regression model is trained in order to extract the feature importance scores for each of the weather parameters in LTW_{s1}. In this case, the eXtreme Gradient Boosting (XGBoost) regression model is selected due to its reduced computational time and regularization parameters which help reduce overfitting. As presented in Table 2.7, LTW_{s1} and Feature A are used as the training input datasets to train the regression model. The total hourly energy demand (E_{total}) for all 20 years of weather data is extracted from the medium office simulation results and used as the training output dataset for the XGBoost regression model.

Table 2.7: Stage 2 XGBoost regression model training datasets

Input Training Dataset	***Output Training Dataset***
DBT	Total energy demand (E_{total})
RH	
DNI	
DHI	
WS	
Feature A	

The decision weather parameters are selected using the extracted feature importance scores as described in step 5 of section 2.3.2. Thus, the monthly feature importance scores for each weather parameter are compared to the monthly feature importance scores obtained for Feature A. As presented in Table 2.8, all five weather parameters considered in Stage 2 are retained. Although no additional filtering occurred due to the use of Feature A in this case, different applications may yield different results. Therefore, the use of Feature A for filtering is recommended to ensure only relevant features are selected.

Table 2.8: Retained weather parameters for Stages 1 and 2

Weather Parameter	***Abbreviation***	***Stage 1***	***Stage 2***
Dry-bulb temperature	DBT	✓	✓
Relative humidity	RH	✓	✓
Direct normal irradiance	DNI	✓	✓
Diffuse horizontal irradiance	DHI	✓	✓
Wind Speed	WS	✓	✓

A subset of the CWEEDS data is created using the five retained features (LTW_{s2}) and used as the training input dataset for Stage 3.

2.4.3 Stage 3: Energy Demand-based Monthly Weighting Factors

The XGBoost regression model is retrained with the retained weather parameters to obtain the undiluted feature importance scores. The CWEEDS subset obtained in Stage 2 (LTW_{s2}) is used as the training input dataset for both models. However, as presented in Table 2.9 and outlined in step 6 of section 2.3.3, in Stage 3, the XGBoost regression model is trained using two different output datasets. In other words, the building energy demand is separated into heating (E_{heat}) and cooling (E_{cool}) demand, and each used as the output training dataset for one of the models.

Table 2.9: Stage 3 XGBoost regression model training dataset

	Input Training Dataset		Output Training Dataset
XGB_{heat}	*LTW_{s2}*	DBT RH DNI DHI WS	Heating energy demand (E_{heat})
XGB_{cool}	*LTW_{s2}*	DBT RH DNI DHI WS	Cooling energy demand (E_{cool})

Following this step, the monthly feature importance scores are obtained for both the XGB_{heat} and XGB_{cool} models for each of the five weather parameters. Prior to selecting the final set of feature importance scores to create the weighting factor file, the average monthly building energy demand is investigated. The final classification attributed to each month based on the procedure outlined in step 7 of section 2.3.3 is presented in Table 2.10.

Table 2.10: Monthly dominant demand type classification

Month	***Heating Dominant***	***Cooling Dominant***
January	✓	
February	✓	
March	✓	
April	✓	
May		✓
June		✓
July		✓
August		✓
September		✓
October	✓	
November	✓	
December	✓	

For each month, the feature importance score corresponding to the dominant demand type, outlined in Table 2.10 are extracted for each parameter. The extracted feature importance scores are concatenated into a single monthly weighting factor file. The monthly weighting factors obtained in this case are presented in Table 2.11. The weighting factor values are further discussed in section 2.5.1 of the results and discussion.

Table 2.11: Monthly weighting factors for generation of proposed TMY weather file

Weather Parameter	***Jan***	***Feb***	***Mar***	***Apr***	***May***	***Jun***	***Jul***	***Aug***	***Sept***	***Oct***	***Nov***	***Dec***
DBT	34.63	30.54	30.53	23.63	45.03	24.57	17.68	17.99	34.82	43.32	25.51	34.75
RH	12.95	11.35	10.30	11.00	10.82	10.26	11.09	10.35	11.05	11.29	13.80	13.21
DNI	9.78	10.10	9.51	16.52	10.27	11.69	13.44	11.14	13.07	13.29	11.21	9.64
DHI	29.66	36.98	40.09	39.20	24.28	43.46	46.05	51.02	31.48	17.34	36.38	29.17
WS	12.97	11.03	9.57	9.64	9.61	10.01	11.73	9.50	9.58	14.76	13.10	13.23

2.4.4 Stage 4: Generation of Proposed TMY Weather File

The monthly weighting factors presented in Table 2.11 are integrated in step iii of the Sandia method to generate the proposed TMY weather file. From the Sandia method, a set of 12 candidate months are selected to create the proposed TMY weather file. The 12 candidate months are extracted from the CWEEDS dataset and condensed into a single EPW file to perform building energy simulations. The output file includes all weather parameters found in CWEEDS, including weather parameters not currently used by EnergyPlus. The selected candidate months for the proposed TMY weather file are presented in Table 2.12.

Table 2.12: Proposed TMY weather file selected candidate months for Montreal

Month	Proposed TMY
January	1999
February	2000
March	2011
April	2013
May	2014
June	2017
July	2015
August	2011
September	2001
October	2000
November	2016
December	2016

2.4.5 Performance Indicator

The root-mean-square error (RMSE) is used to evaluate and compare the performance of the proposed TMY weather file to CWEC. As previously stated, TMY weather files are single-year files intended to represent the average long-term weather data. Therefore, the average monthly building energy demand, obtained from the long-term building energy simulation results, is used as the baseline for comparison. The RMSE is calculated using the monthly energy demand for both heating and cooling as outlined in equation ((2.3).

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(TMY_i - LTA_i)^2}{N}} \tag{2.3}$$

TMY: Monthly heating or cooling demand results simulated using the typical meteorological year weather file.

LTA: Average monthly heating or cooling demand simulated using the long-term weather data.

N: Total number of months in a year

The average monthly energy demand is used as the baseline to reduce the impact of outliers within the long-term weather dataset, which may skew the results, as RMSE penalizes larger errors.

Additionally, although employed in many studies, the use of annual energy demand as a baseline for comparison does not provide an adequate picture of the TMY file's overall performance. Although the annual energy demand may closely reflect that of the long-term average, the possibility remains that the monthly demand results may exhibit significant deviations (under/overestimation).

2.5 Results and Discussion

The results and discussion section is divided into two main parts. First, an initial review of the generated proposed TMY weather file is presented in section 2.5.1. The proposed TMY file is then evaluated and compared to CWEC using building energy demand in section 2.5.2.

2.5.1 Proposed Typical Meteorological Year Weather File Review

2.5.1.1 Monthly weighting factors

In Stage 3, the monthly weighting factors are extracted and used to generate the proposed TMY weather file. As previously presented in Table 2.11, the weighting factors obtained for the decision weather parameters vary from month to month. Dry-bulb temperature is considered to have the greatest impact on building energy demand, followed by the combined solar irradiance parameters. As demonstrated in Figure 2.6, notable fluctuations in weighting factors occur between March and May, as well as between September and November. These significant changes are likely due to the seasonal shift in weather which occur in spring and autumn. These periods are referred to as shoulder seasons and often exhibit high fluctuations in temperature. Therefore, the variation in weight attributed to dry-bulb temperature is likely associated to the fluctuation in temperature occurring during both these periods.

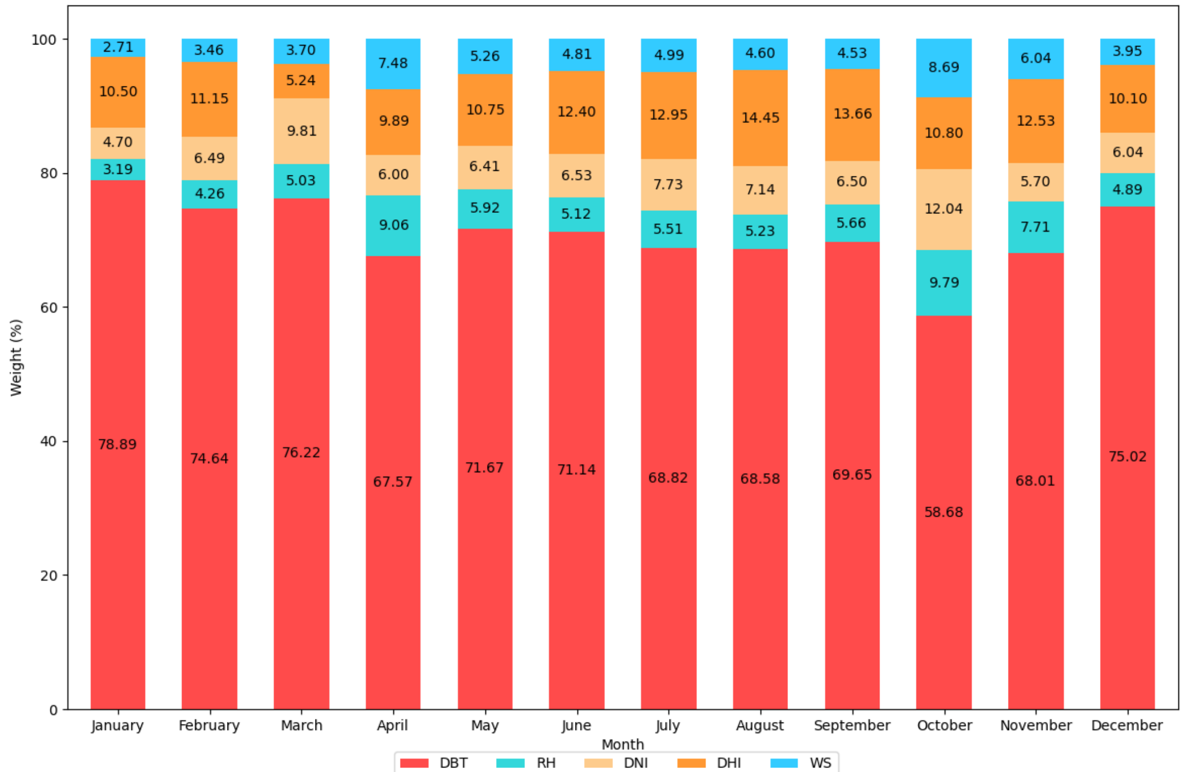

Figure 2.6: Monthly weighting factors for each decision weather parameter.

The impact of seasonal weather fluctuations on the weighting factors is further supported by the resulting weights extracted for the remaining weather parameters, which demonstrate significant fluctuations during the shoulder months. The jump in importance for direct normal irradiance relative to diffused horizontal irradiance, which occurs in March and October, is likely the result of these seasonal shifts. The fluctuations in weight occurring during the shoulder seasons demonstrates the need for seasonal weighting factors. However, further investigation into the impact of seasonal weather variations on building energy performance is required.

2.5.1.2 Selected years

As previously discussed, the proposed TMY file is generated following the steps outlined in the Sandia method, using CWEEDS. The purpose of using both the Sandia method and CWEEDS to generate the weather file is to facilitate the comparison between CWEC and the proposed TMY. The selected candidate months for both CWEC and the proposed TMY weather file are presented in Table 2.13.

Table 2.13: CWEC and proposed TMY weather file selected candidate months for Montreal

Month	*CWEC*	*Proposed TMY*
January	1999	1999
February	2009	2000
March	2009	2002
April	2013	2013
May	2014	2014
June	2017	2008
July	2013	2015
August	1998	1998
September	2001	2001
October	2000	2000
November	2016	2005
December	2003	2016

Although both files have significant differences in weighting factors and decision weather parameters, in many cases, the same candidate month is selected. This is due to both the weather data distribution within the file, as well as the Sandia method selection process. In the Sandia method, the top five candidate months are selected based on the weighted sum, which considers the weighting factors attributed to each weather parameter. However, following the weighted sum calculation, the candidate months are reranked based on the mean and median error for dry-bulb temperature and global horizontal irradiance. The candidate months are then evaluated in succession using the persistence criteria until a month is selected. Therefore, although the initial ranking and selection of the top five candidate months are influenced by the weighting factors, the final candidate month is selected based on the results of the subsequent statistical tests. Further investigation is required to ascertain the impact of the persistence criteria on the resulting weather files.

2.5.2 Energy Demand Comparison

In this section, the energy demand simulation results for both CWEC and the proposed TMY file are evaluated and compared to the CWEEDS long-term average (LTA) simulation results.

2.5.2.1 Annual energy demand

In Figure 2.7, the annual energy demand results obtained for CWEC, the proposed TMY and the LTA are compared to the CWEEDS simulation results. The difference in annual energy demand is obtained for all three cases and presented in the figure.

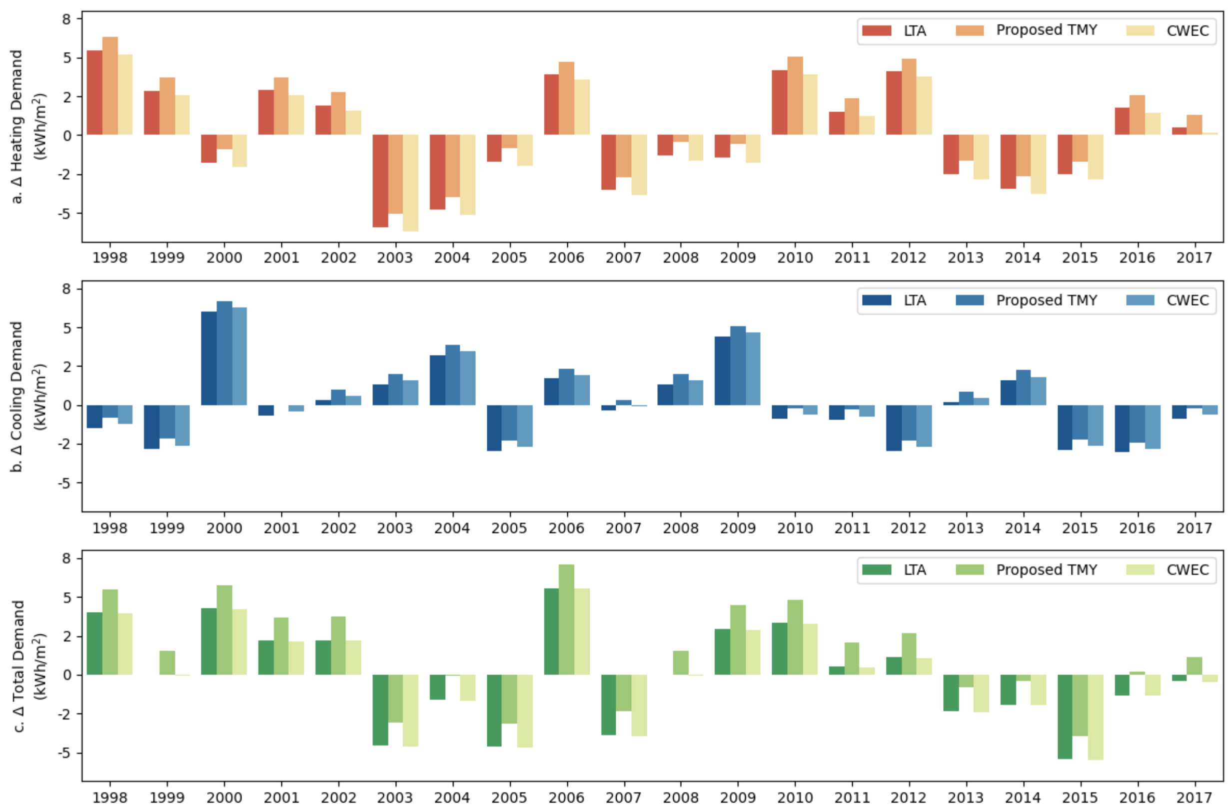

Figure 2.7: Comparison of annual energy demand results for reference year weather files and CWEEDS dataset.

The importance in using the LTA to evaluate the reference year files stems from the large variations in energy demand which may occur on a year-by-year basis, as is demonstrated in Figure 2.7. As demonstrated in the figure, the heating and cooling energy demand obtained using CWEC appear to better reflect the long-term average. However, as previously discussed in section 0, the annual energy demand may not be an adequate baseline for comparison, as annual energy demand may mask significant deviations in the monthly energy demand.

2.5.2.2 Monthly energy demand

The monthly energy demand for each case is presented in Figure 2.8 for both heating and cooling. As presented in Table 2.13, CWEC and the proposed TMY weather file have different candidate months for February, March, June, July, November, and December. In other words, the resulting monthly demand for both files will be relatively identical, except for the abovementioned six months. The difference in heating demand between the long-term average and the reference year weather files for February and December are quite similar. In March, a noticeable discrepancy in heating demand to the LTA is noted for CWEC, contrary to the heating demand obtained using the proposed TMY weather file. In the case of cooling demand, the proposed TMY results are closer to the LTA compared to the CWEC results for all months except June. A table of the monthly energy demand results for all three cases are presented in Appendix A.

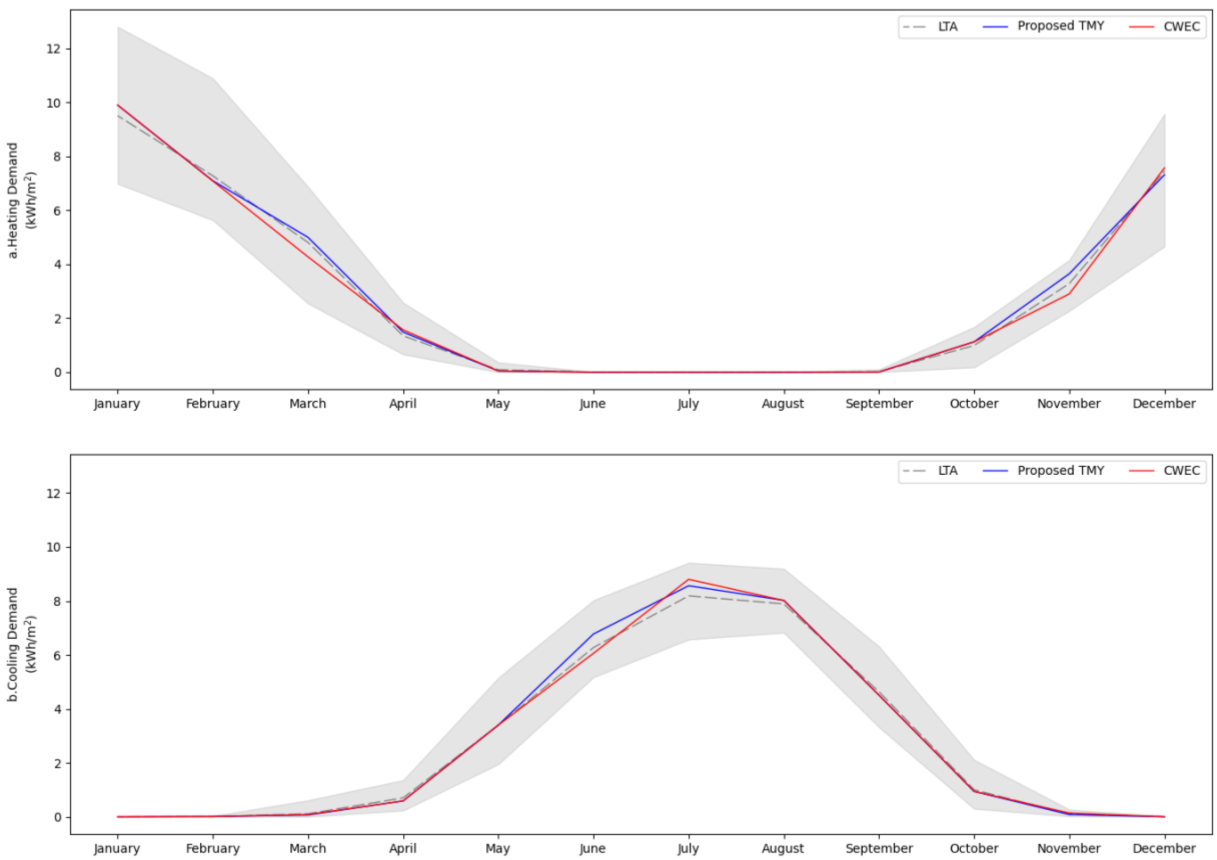

Figure 2.8: Monthly energy demand comparison for a. heating and b. cooling

The results presented in Figure 2.8 emphasize the impact of solely relying on total annual energy demand to evaluate typical meteorological year weather files. In this case, the CWEC total annual energy demand results, presented in Figure 2.7, showed strong similarities to the LTA results. However, as shown in Figure 2.8 the discrepancy in monthly demand appears to be greater for CWEC than the proposed TMY weather file.

In order to quantify the discrepancy presented in Figure 2.8, the root-mean-square error is obtained following the procedure outlined in section 0 and presented in Table 2.14. To help interpret and compare the RMSE values obtained, a RMSE reduction value is included within the table. The RMSE reduction is based on the percentage of relative change formula and uses the CWEC RMSE as the baseline. In this case, a negative value indicates a reduction in the RMSE, whereas a positive value indicates an increase in error.

Table 2.14: Root-mean-square error obtained for CWEC and the proposed TMY for monthly heating and cooling demand.

Energy Demand Type	***CWEC***	***Proposed TMY***	***RMSE Reduction***
Heating	1 675.28	1 278.98	***-23.65 %***
Cooling	1 369.55	1 329.45	***-2.93 %***
Total	2 100.94	1 763.76	***-16.05 %***

The proposed TMY weather file yields simulation results with a smaller RMSE than CWEC for both monthly heating and cooling demand. The RMSE reduction values show a decrease in RMSE of 16.05% for total building energy demand.

These results further highlight the inadequacy in using annual energy demand as the sole means to evaluate typical meteorological year weather files initially emphasized in Figure 2.8. The decrease in error is likely attributed to the use of monthly weighting factors, which allow to account for the impact of seasonal changes on building energy demand.

2.5.3 Limitations

In this study, a ML regression model is employed to define the generation parameters in order to provide an adaptable approach. However, key limitations of the proposed framework stem from the ML regression model training requirements, which impose certain constraints on the training datasets.

The first limitation arises from the ML regression model accuracy constraints, which often require the exclusion of certain variables within the training datasets. The current model requires variables with continuous data to preserve the performance. Although this issue is not prevalent for the current case study, the constraint may pose an issue when investigating different applications requiring discrete or categorical variables. An example of this issue occurs for natural ventilation, which is significantly impacted by wind direction, a discrete variable. Therefore, although the proposed framework may be used to define the generation parameters, the resulting weather files may not be optimized for the given application.

Furthermore, the current framework is location dependent, as no generalized approach has been developed to group training datasets by climate. As such, climate-based investigations may require further research to adapt the current framework into a generalized approach.

2.5.4 Future Research

The primary contribution of this study is the introduction of a data-driven framework to define generation parameters which may be adapted for different investigations. The proposed approach relies on long-term weather data and simulation results to train the regression model to define the generation parameters. Therefore, the approach may be adapted for a variety of applications and file types by utilizing the corresponding simulation results.

The present framework provides several future research opportunities including the implementation of the current work for varying applications such as hygrothermal analysis and solar energy systems. Further research may be conducted on the implementation of the proposed methodology for alternative reference weather file types, such as extreme weather files, by considering simulation variables in the training process. Furthermore, investigations into the impact of varying climates on the performance of the resulting TMY weather files should be conducted.

Although the current framework may be used for different applications and climates, additional research is required to address the limitations identified in this study. Potential research directions aimed at addressing these constraints include refining the framework to accommodate discrete variables without compromising the model performance, as well as developing a generalized approach suitable for climate-based investigations.

2.6 Conclusion

The current application of the Sandia Laboratory method to generate TMY weather files relies on universal generation parameters. The decision weather parameters, and their corresponding weights, are based on expert judgement and often neglect variations in climate and application. The purpose of the methodology presented in this study is to provide a data-driven approach to define the weather parameters and weighting factors, accounting for seasonal variations. Furthermore, the proposed methodology aims to provide a framework suitable for varying applications and for specific weather data limitation cases.

The proposed methodology is divided into a four-stage process and adopts a gradient boosted tree regression model to define the generation parameters used in the Sandia method to construct the resulting TMY weather files. The regression model is used to determine the decision weather parameters and their corresponding weights by extracting the feature importance for each of the considered parameters. The TMY weather file is generated using a set of monthly weighting factors obtained for each of the selected decision weather parameters.

The proposed approach was used to generate a TMY weather file for Montreal using the building energy demand of a prototypical medium office building. The resulting decision weather parameters obtained in the case study differed from those used to generate CWEC. Dry-bulb temperature and diffuse horizontal irradiance were found to have the greatest impact on energy demand. Furthermore, the resulting monthly weighting factors obtained for both parameters exhibited inversely proportional seasonal fluctuations.

The TMY weather file generated using the defined generation parameters obtained from the presented methodology was evaluated against CWEC using the LTA building energy demand. The proposed TMY weather file demonstrated a good agreement with the LTA annual heating energy demand. Whereas CWEC was found to better demonstrate the LTA annual cooling demand.

However, when considering monthly energy demand, the TMY weather file outperformed the CWEC file for both heating and cooling with an improvement in RMSE of 23.65% and 2.93% respectively. These results correspond to a 16.04% improvement when considering total energy demand. These results demonstrate the inadequacy of universal annual weighting which fail to account for seasonal variations. Although annual energy demand may demonstrate good agreement, the use of annual weighting factors may lead to significant discrepancies in monthly energy demand.

Presently, the procedures used to generate most of the widely available TMY weather files do not account for variations in climate and their associated seasonal changes. The presented methodology seeks to address this issue by considering the impact of the local weather data on energy demand to define the generation parameters. The presented approach offers a certain flexibility regarding the training data and may be used in instances with limited weather data or adapted for different applications. Further investigation is needed to ascertain the impact of varying climates and applications on the resulting TMY weather file using the abovementioned approach. Finally, additional research may be pursued to adapt the current methodology to address model constraints and to develop a more generalized approach.

2.7 Acknowledgement

The authors acknowledge that this work was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) [RGPIN-2016-05711].

2.8 References

[1] R. Morris, "Final Report - Updating CWEEDS Weather Files," Environment and Climate Change Canada, 2016.

[2] H. Lund, "The Design Reference Year Users Manual," Thermal Insulation Laboratory, Technical University of Danmark, 1995.

[3] R. Festa and C.F. Ratto, "Proposal of a numerical procedure to select Reference Years," Sol. Energy, vol. 50, no. 1, pp. 9-17, 1993.

[4] Hall, R. Prairie, H. Anderson and E. Boes, "Generation of Typical Meteorological Years for 26 SOLMET Stations," Sandia National Laboratories, Albuquerque, 1978.

[5] National Climatic Center, “Typical Meteorological Year User’s Manual”, 1981.

[6] International Organization for Standardization, "ISO Standard 15927-4, Hygrothermal performance of buildings - Calculation and presentation of climatic data - Part 4: Hourly data for assessing the annual energy use for heating and cooling.," 2005.

[7] K. Skeiker, "Comparison of methodologies for TMY generation using 10 years data for Damascus, Syria," Energy Conversion and Management, vol. 48, no. 7, pp. 2090-2102, 2007.

[8] D. B. Crawley, "Which weather data should you use for energy simulations of commercial buildings?," ASHRAE Transactions, vol. 104, pp. 498-515, 1998.

[9] Y. Jiang, "Generation of typical meteorological year for different climates of China," Energy, vol. 35, p. 1946–1953, 2010.

[10] D. Seo, Y. J. Huang and M. Krarti, "Impact of Typical Weather Year Selection Approaches on Energy Analysis of Buildings," ASHRAE Transactions, vol. 116, no. 1, pp. 416-427, 2010.

[11] W. Marion and K. Urban, " User's Manual for TMY2s (Typical Meteorological Years) - Derived from the 1961-1990 National Solar Radiation Data Base," National Renewable Energy Laboratory , Golden, 1995.

[12] S. Wilcox and W. Marion, "Users Manual for TMY3 Data Sets," National Renewable Energy Laboratory, Golden, 2008.

[13] D. J. Thevenard and A. P. Brunger, "The development of typical weather years for international locations: Part I, Algorithms," ASHRAE Transactions, vol. 108, p. 376, 2002.

[14] Y. J. Huang, F. Su, D. Seo, and M. Krarti, "Development of 3012 IWEC2 Weather Files for International Locations (RP-1477)," ASHRAE Transactions, vol 120, no. 1, 2014.

[15] Environment and Climate Change Canada, "Canadian Weather Energy and Engineering Datasets (CWEEDS files) and Canadian Weather Year for Energy Calculations (CWEC Files) User's Manual," 2008.

[16] A.L.S. Chan, "Generation of typical meteorological years using genetic algorithm for different energy systems," Renewable Energy, vol. 90, pp. 1-13, 2016.

[17] G. Georgiou, M. Eftekhari, P. Eames and M. Mourshed, "A study of the effect of weighting indices for the development of TMY used for building simulation," in 13th Conference of International Building Performance Simulation Association, Chambery, 2013.

[18] H.D. Kambezidis, B.E. Psiloglou, D.G. Kaskaoutis, D. Karagiannis, K. Petrinoli, A. Gavriil and K. Kvadias, "Generation of typical meteorological years for 33 locations in Greece;

Adaptation to the needs of various applications," Theoretical and Applied Climatology, vol. 141, no. 3-4, pp. 1313-1330, 2020.

[19] H. Yang and L. Lu, "Study of Typical Meteorological Years and Their Effect on Building Energy and Renewable Energy Simulations," ASHRAE Transactions, vol. 110, pp. 424-431, 2004.

[20] J. Sun, J. Li, F. Xiao and J. Xiao, "Generation of typical meteorological year for integrated climate based daylight modeling and building energy simulation," Renewable Energy, vol. 160, pp. 721-729, 2020.

[21] F. Meng, M. Li, J. Cao, J. Li, M. Xiong, X. Feng and G. Ren, "The effects of climate change on heating energy consumption of office buildings in different climate zones in China," Theoretical and Applied Climatology, vol. 133, no. 1-2, pp. 521-530, 2017.

[22] T. Kalamees, K. Jylhä, H. Tietäväinen, J. Jokisalo, S. Ilomets, R. Hyvönen and S. Saku, "Development of weighting factors for climate variables for selecting the energy reference year according to the EN ISO 15927-4 standard," Energy and Buildings, vol. 47, pp. 53-60, 2012.

[23] T. Hong, W. K. Chang and H. W. Lin, "A fresh look at weather impact on peak electricity demand and energy use of buildings using 30-year actual weather data," Applied Energy, vol. 111, pp. 333-350, 2013.

[24] H. Li, J. Huang, Y. Hu, S. Wang, J. Liu and L. Yang, "A new TMY generation method based on the entropy-based TOPSIS theory for different climatic zones in China," Energy, vol. 231, p. 120723, 2021.

[25] S. Gai, X. Zhang, J. Xie and K. Yin, "A Method for Generating Typical Meteorological Year with Simplified Types of Meteorological Parameters," presented at 5th Int. Conf. on Building Energy and Environment (COBEE), Montreal, 2022.

[26] O.S. Ohunakin, M.S. Adaramola, O.M. Oyewola and R.O. Fagbenle, "Generation of a typical meteorological year for north–east, Nigeria," Applied Energy, vol. 112, pp. 152-159, 2013.

[27] H. Zang, M. Wang, J. Huang, Z. Wei and G. Sun, "A Hybrid Method for Generation of Typical Meteorological Years for Different Climates of China," Energies, vol. 9, p. 1094, 2021.

[28] M. Kuhn and K. Johnson, Feature Engineering and Selection: A Practical Approach for Predictive Models, 1st ed., Boca Raton, Florida: Chapman and Hall/CRC, 2019.

[29] T. Hastie, R. Tibshirani and J. Friedman, The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Stanford: Springer, 2008.

[30] M. Hosseini, A. Bigtashi and B. Lee, "A systematic approach in constructing typical meteorological year weather files using machine learning," Energy & Buildings, vol. 226, no. 110375, 2020.

[31] U.S. Department of Energy, “Commercial reference buildings,” Energy.gov, https://www.energy.gov/eere/buildings/commercial-reference-buildings (accessed Oct. 20, 2023).

[32] NECB, "National Energy Code of Canada for Buildings 2020," National Research Council of Canada, 2022.

Chapter 3 Evaluating the applicability of a machine learning methodology to improve TMY weather file generation for different Canadian climate zones.

3.1 Contribution of Authors

This manuscript has been submitted and is currently under review with the Journal of Building Engineering. The journal paper demonstrates the applicability of the methodology proposed in Chapter 2 to varying climates and proposes the use of a standardized set of climate zone-based weighting factors to allow for greater adoption of the proposed methodology. Ashleigh Papakyriakou is the first author of the paper. Ashleigh's contributions include the conceptualization, development and validation of the energy models, data curation, formal analysis, writing, editing and review. Anahita Bigtashi (second author) contributions include the conceptualization, development of the proposed weather files ($TMY_{STATION}$ and TMY_{CZ}), data curation, editing and review. Dr. Bruno Lee contributed to conceptualization, and review.

3.2 Introduction

Building energy modelling is used for many purposes such as meeting code compliance, or as a design tool to help with decision-making to create more sustainable buildings. Typically, building energy models use typical metrological year (TMY) weather files to reduce the computational time associated with multi-year simulations. TMY weather files allow designers to be more efficient as they can run a single simulation instead of multiple simulations to illustrate how the building will perform over the long term. A TMY weather file is a single artificial year that is composed of typical months selected from historical data over a specified period to represent the long-term average weather data. The accuracy of the results of the simulation depends on how well the TMY weather file can represent the long-term average weather data. Canadian cities with sustainable building requirements for mid-to-high rise, commercial and mixed-use buildings such as the City of Vancouver [1], and the City of Toronto [2] require the use of Canadian Weather Year for Energy Calculation (CWEC) [3] TMY weather files for building simulation to verify code compliance.

Various organizations have developed TMY weather files which are commonly used for building energy modelling. These datasets include the CWEC [3], the International Weather for Energy

Calculation (IWEC [4] and IWEC2 [5]), as well as the Typical Metrological Year ([TMY][2] [6], TMY2 [7] and TMY3 [8]). CWEC files are created by Environment and Climate Change Canada, focusing on Canadian locations, IWEC, and IWEC2 files are produced by ASHRAE for locations outside of Canada and the United States, and [TMY][1] files are produced by NREL for locations in the United States. TMY2 and TMY3 are updated versions of the [TMY][1] file format. All these datasets use the Sandia method to generate the TMY weather files. Common TMY weather file datasets use different weighting factors to determine the typical months. Furthermore, the TMY2 and TMY3 datasets include direct normal irradiance as a parameter. The weighting factors used to generate common TMY weather files have been assigned subjectively by an expert based on the intended use of the weather file [9].

In a previous study, the authors [10] extensively reviewed the influence of weighting factors on the generation of TMY weather files, the key findings from the literature review are summarized as follows. Georgiou et al. [11] conducted a study highlighting the impact the weighting factors have on the typical months selected to create the TMY weather file. Since building heating and cooling demand is greatly influenced by weather parameters, selecting different months to create the TMY weather file will impact the results. Kalamees et al. [12] highlighted the seasonal variation in the influence of the weather parameters, demonstrating the need for monthly weighting factors. Hong et al. [13] found large discrepancies, particularly in colder climates, where TMY3 results often under- or overestimated energy savings compared to the historical weather data results. The study demonstrated the need for developing weighting factors that consider local climate conditions. Meng et al. [14] found dry bulb temperature to have the largest influence on heating demand across the three climate zones however the magnitude of influence varied based on the climate zone as well as the influence from the other weather parameters. The study highlighted how the influence of weather parameters varies based on the climate zone. The studies demonstrate the need for monthly customized weighting factors that consider local climate conditions.

[2] *[TMY]: Used to denote the Typical Meteorological Year file format originally presented in.*

The use of universal weather parameters and weighting factors fail to account for variations in climate which can lead to significant discrepancies in the simulation results. To address this problem, Bigtashi et al. [10] proposed a methodology using machine learning to systematically determine relevant weather parameters and weighting factors based on local climate conditions to improve TMY weather file generation. The methodology used a machine learning regression model to select weather parameters and generate their respective weighting factors using feature importance. The methodology by Bigtashi et al. [10] is divided into four stages: (1) the methodology uses long-term historical weather data and building simulation to obtain hourly energy demands; (2) the energy demands are subsequently used to train the machine learning regression model alongside the long-term weather data; (3) the relevant weather parameters and weighting factors are extracted using feature importance; (4) the weather parameters and their corresponding weights are integrated into the Sandia method to generate the TMY weather files. The approach was applied to a medium office building in Montreal, Canada. The results showed an improvement in the representativeness of the long-term average building energy demand for the weather file generated with the proposed methodology compared to the commonly used Canadian TMY weather file, CWEC. The monthly RMSE values for the total building energy demand indicated the proposed TMY weather file outperformed the CWEC weather file by 16%. Although the results demonstrated a significant improvement, a notable constraint outlined by the authors is the location-dependent nature of the methodology. The methodology requires customized weighting factors to be generated for each TMY weather file location. This process can be very time-consuming and potentially limit the adoption of the proposed methodology.

A few studies [13] [14] have emphasized the need for weighting factors to account for local climate conditions. Widely recognized building energy standards, such as NECB [15], ASHRAE 90.1 [16], and CaGBC Zero Carbon Building Design Standard [17], organize design requirements based on climate zones. Recognizing the time-intensive nature of generating location-dependent weighting factors, there exists a potential strategy to group these factors by climate zones, maintaining consideration for local climate conditions. The adoption of climate zone-based weighting factors, as opposed to location-dependent ones, presents an opportunity to streamline the process of generating TMY weather files while still capturing the nuanced effects of local climates.

The methodology proposed by Bigtashi et al. [10] demonstrated a significant improvement in weather file performance compared to the CWEC weather files, however, the study was limited to one location and the methodology requires location-dependent weighting factors which can be very time-consuming to generate for multiple locations. To address these limitations, the study has outlined two main objectives. The objectives of this study are:

1. To assess the applicability of the methodology proposed by Bigtashi et al. [10] across varying Canadian climates,
2. And to investigate the feasibility of using standardized climate zone-based weighting factors to reduce the computational time associated with extracting location-based weighting factors while still considering local climate conditions to facilitate wider adoption of the proposed methodology.

3.3 Methodology

The purpose of this study is to evaluate the applicability of Bigtashi et al.'s [10] methodology for different climate zones across Canada and to examine the feasibility of employing standardized climate zone-based weighting factors to produce TMY weather files. The adoption of climate zone-based weighting factors streamlines the process and reduces the time required by eliminating the need to create location-specific weighting factors.

To investigate the applicability of the methodology by Bigtashi et al. [10] and the feasibility of the standardized climate zone-based weighting factors, a case study is conducted using a prototypical medium-sized office building for 18 cities across the six different Canadian climate zones. Canada is selected for the case study due to the diverse range of climate zones, varying from mild to very cold. The performance of the TMY weather files is assessed using the hourly long-term building energy demand of a medium prototypical office building.

The methodology is divided into five main sections which are displayed in Figure 3.1 along with a flow chart of the process. Section 3.3.1 provides an overview of the 18 Canadian cities and the corresponding hourly long-term historical weather data for each location. Section 3.1.2 reviews the six building energy models developed for the study which follow the National Energy Code of Canada for Buildings (NECB) requirements. Section 3.3.2 discusses the generation methods used to produce the two TMY weather files for each location. The TMY weather files generated with the location-based weighting factors are denoted as the $TMY_{STATION}$ weather files, and the TMY weather files generated with the climate zone-based weighting factors are denoted as TMY_{CZ} weather files. The total number of simulations are outlined in section 3.3.3. Lastly, in section 3.3.4, the performance indicators used in this study are presented. The NMBE and the CV(RMSE) values for both the proposed TMY weather files and CWEC weather files will be calculated with the long-term weather data to evaluate the results.

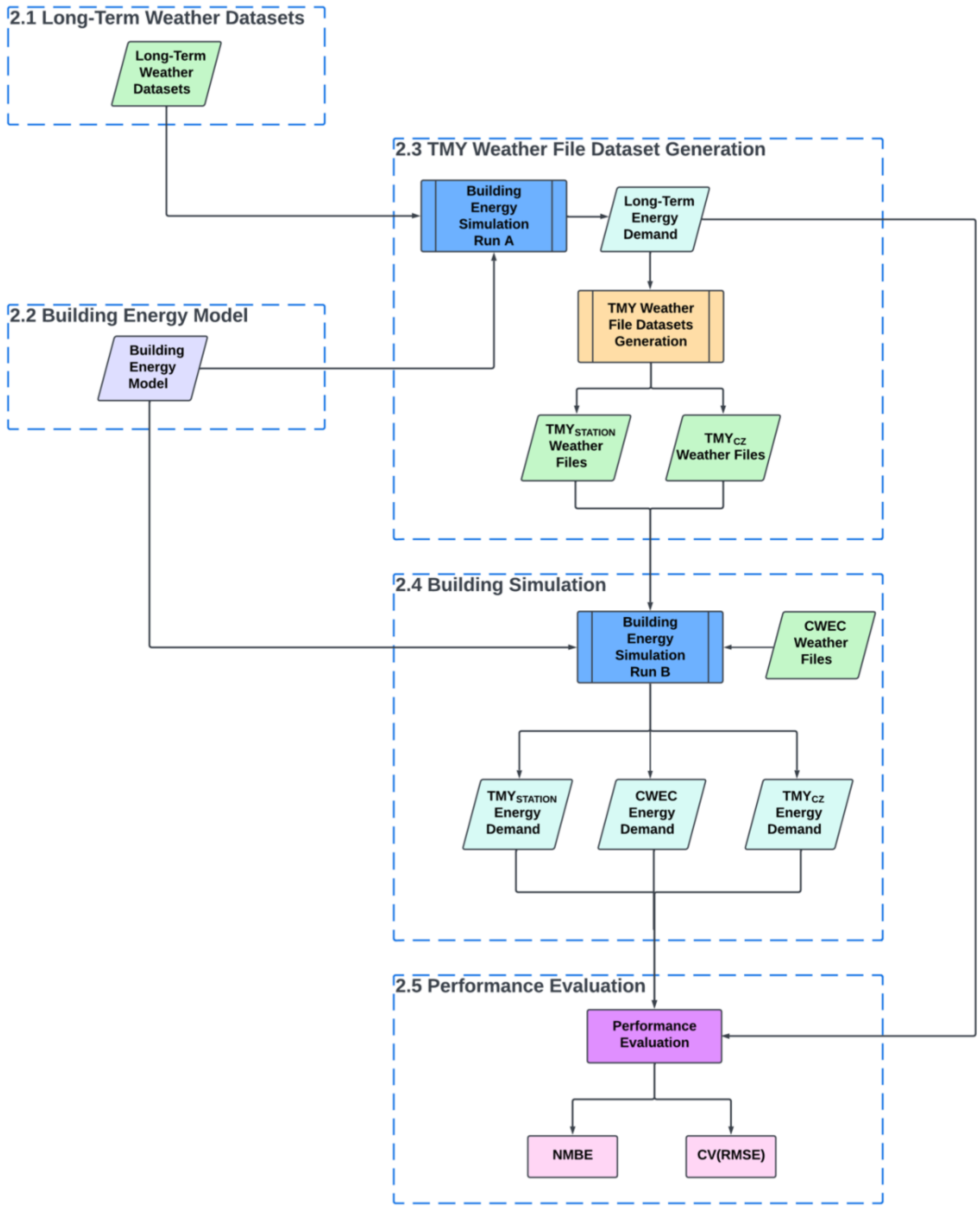

Figure 3.1: Case study methodology

3.3.1 Long-Term Weather Datasets

To generate the TMYSTATION and TMYCZ weather files, hourly long-term weather (LTW) data is acquired for different cities across Canada. Environment and Climate Change Canada provides the

Canadian Weather Energy and Engineering Datasets (CWEEDS) files for 564 Canadian locations. These files contain hourly weather data for at least 10 years, ranging between 1998 to 2017, and contain the weather parameters required for the building energy simulation software, Energy Plus [18]. The CWEEDS files are also used to generate the CWEC weather files using the Sandia method.

There are six different climate zones in Canada which are defined based on heating degree days (HDD) at 18°C [15]. The six climate zones and their corresponding HDD ranges are presented in Table 3.1.

Table 3.1: Canadian climate zones defined by HDD [15]

	Climate Zone					
	4	5	6	7A	7B	8
HDD	< 3000	3000 to 3999	4000 to 4999	5000 to 5999	6000 to 6999	≥7000

In this study, three cities are selected for each Canadian climate zone, resulting in a total of 18 different cities. The CWEEDS file was obtained for each Canadian city as shown in Figure 3.2. The city markers are colour-coded based on their respective climate zone.

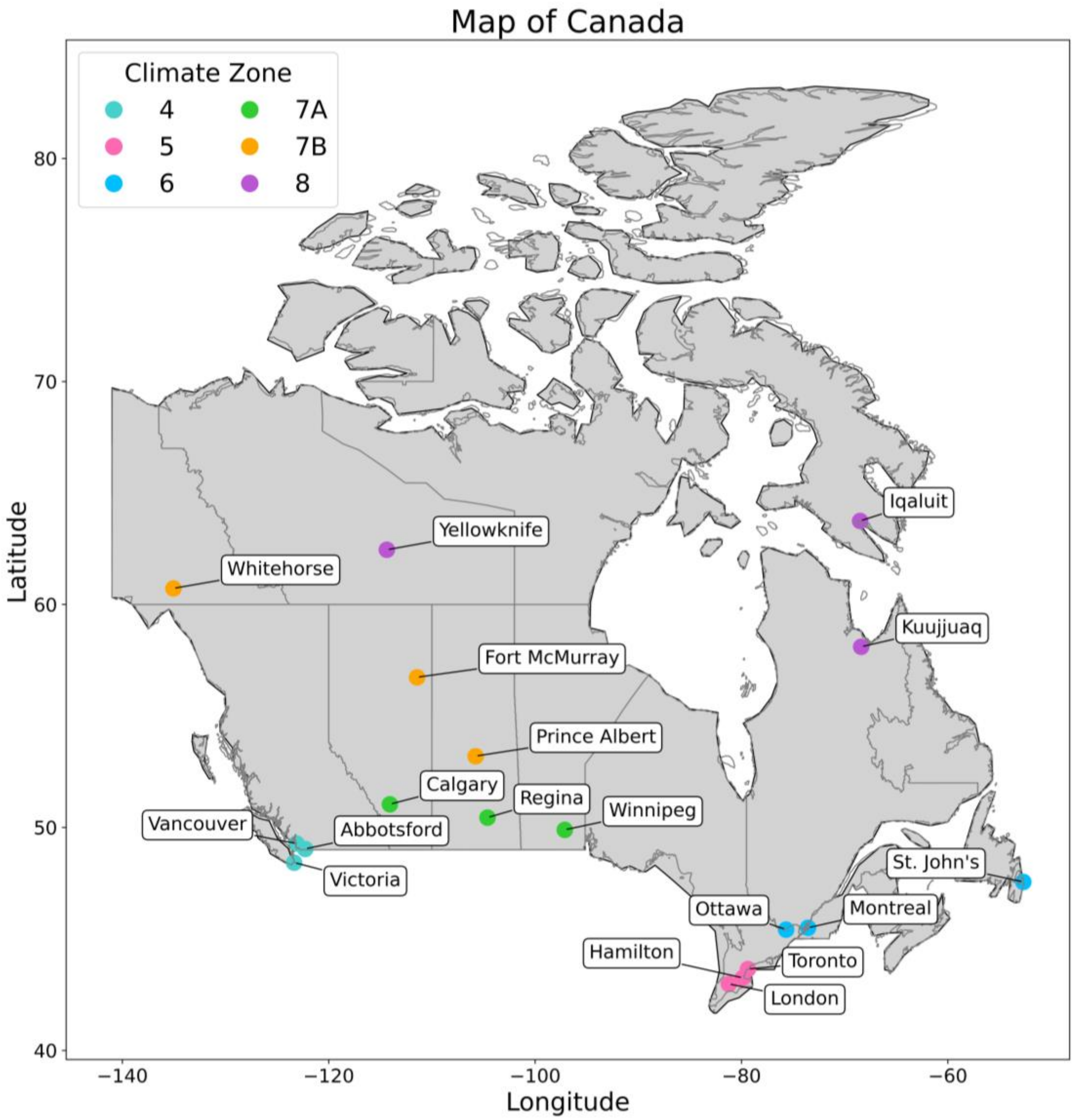

Figure 3.2: Map of the cities selected for weather file generation.

Table 3.2 shows each city, and its respective climate zone, heating degree days, weather station, weather station name and the period of data available. The airport weather station was chosen to represent each city for consistency, as some locations only had the airport weather station available.

Table 3.2: Case study cities

Climate Zone	City	Province	HDD [30]	Station ID	Station Name	Period Range
4	Vancouver	British Columbia	2825	1108395	Vancouver Intl. Airport	1998-2017
	Victoria	British Columbia	2650	1018621	Victoria Intl. Airport	1998-2017
	Abbotsford	British Columbia	2860	1100031	Abbotsford Airport	1998-2017
5	Toronto	Ontario	3520	6158731	Toronto Intl. Airport	1998-2017
	Hamilton	Ontario	3460	6153193	Hamilton Airport	2000-2017
	London	Ontario	3900	6144473	London Airport	1998-2017
6	Ottawa	Ontario	4500	6106001	Ottawa Intl. Airport	1998-2017
	Montreal	Quebec	4200	7025251	Montreal Intl. Airport	1998-2017
	St. John's	Newfoundland and Labrador	4800	8403505	St. Johns Intl. Airport	1998-2017
7A	Calgary	Alberta	5000	3031092	Calgary Intl. Airport	1998-2017
	Regina	Saskatchewan	5600	4016566	Regina Intl. Airport	1998-2017
	Winnipeg	Manitoba	5670	5023227	Winnipeg Intl. Airport	1998-2017
7B	Whitehorse	Yukon	6580	2101303	Whitehorse Airport	2005-2017
	Fort McMurray	Alberta	6250	3062697	Fort McMurray Airport	1998-2017
	Prince Albert	Saskatchewan	6100	4056241	Prince Albert Airport	1998-2017
8	Kuujjuaq	Quebec	8550	7113535	Kuujjuaq Airport	2005-2017
	Iqaluit	Nunavut	9980	2402596	Iqaluit Airport	2005-2017
	Yellowknife	Northwest Territories	8170	2204101	Yellowknife Airport	2005-2017

The data from the CWEEDS files for the outlined cities above is used in section 3.3.2 to generate the $TMY_{STATION}$ and TMY_{CZ} weather files for each city. The CWEEDS data is also used to evaluate the TMY weather files' performance by comparing the building energy demands.

3.1.2 Building Energy Model

For the case study, six energy models of a typical medium-sized office building are created, one for each Canadian climate zone as discussed in the previous section. The medium office geometry is based on the DOE medium office parameters [19]. The building energy models are used to both generate the TMY weather files as well as to evaluate the weather file performance. The energy model was designed to the National Energy Code of Canada for Buildings (NECB) 2020 [15] code requirements. The internal gains and schedules are consistent between each energy model to facilitate comparison and are shown in Table 3.3. Furthermore, the heating set point is 21°C and the cooling set point is 24°C, which is consistent for all building energy models.

Table 3.3: NECB 2020 Office Internal Gains

Space Type	Lighting [W/m²]	Receptacle Equipment [W/m²]	Occupancy [m²/person]	Schedule [-]	Infiltration [L/s/m² of façade at 5 Pa]
NECB Office	10.00	7.50	25.00	Schedule A	0.25

The envelope requirements outlined in NECB 2020 [15] for each climate zone are shown in Table 3.4. The NECB 2020 [15] slab overall thermal transmittance is converted to the ASHRAE 90.1-2022 [16] F-factor requirements for the energy models.

Table 3.4: Climate zone envelope requirements [15] [16]

Parameter	Climate Zone					
	4	5	6	7A	7B	8
Walls USI [W/m²K]	0.29	0.27	0.24	0.22	0.19	0.17
Roof USI [W/m²K]	0.16	0.16	0.14	0.12	0.12	0.11
Window [W/m²K]	1.9	1.9	1.73	1.73	1.44	1.44
Slab on Grade [W/m²K]	0.757 for 1.2 m	0.757 for 1.2 m	0.757 for 1.2 m	0.757 for 1.2 m	0.757 for 1.2 m	0.38
ASHRAE F-factor [W/m²K]	1.13	1.13	1.13	1.13	1.13	0.52
WWR [-]	0.4	0.4	0.39	0.33	0.23	0.2

The model geometry of the medium office with a window-to-wall ratio (WWR) of 40% is shown in Figure 3.3.a). The energy model geometry shown is used for Climate Zone 4 and Climate Zone 5, the energy models for the remaining climate zones have the same geometry but have a smaller WWR ratio. The office building used for all climate zones consists of three storeys and has a total area of 6,898 m^2. Each level is broken up into five zones, four perimeter zones and one core zone, as shown in Figure 3.3 b).

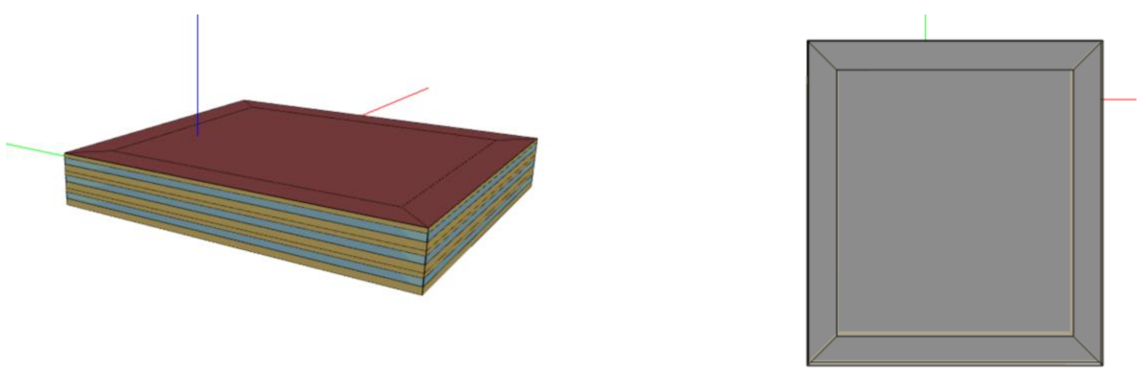

Figure 3.3: a) Model Geometry *b) Model Zoning*

3.3.2 TMY Weather File Dataset Generation

The LTW data obtained in section 3.3.1 along with the building energy models developed in section 3.1.2 are used to determine the weather parameters and respective weighting factors by applying the methodology proposed by Bigtashi et al. [10] for each weather station outlined in section 3.3.1. The location-based weather parameters and weighting factors obtained through the methodology are integrated into the Sandia method to produce a TMY weather file for each location. The TMY weather files generated with the location-based weighting factors for each weather station are referred to as the $TMY_{STATION}$ weather files. A total of 18 $TMY_{STATION}$ weather files will be generated, one for each weather station.

Since the proposed methodology is location-dependent, a standardized set of weighting factors based on the various Canadian climate zones are determined and used to generate an additional TMY weather file for each location. These weighting factors are referred to as the climate zone-based weighting factors. The weather files generated with the climate zone-based weighting factors are referred to as the TMY_{CZ} weather files.

To generate the TMY_{CZ} weather files, first, the climate zone-based weighting factors are obtained by averaging the $TMY_{STATION}$ weighting factors for each weather parameter across the three cities within each designated climate zone, as shown in equation (3.1)

$$WF_{CZ,ij} = Average\ (WF_{ST,j}) \tag{3.1}$$

Where:

j is the weather parameter.

WF_{ST} is the location-based weighting factor.

i is the climate zone.

WF_{CZ} is the climate zone-based weighting factor.

For example, to obtain the Climate Zone 4 dry bulb temperature climate zone-based weighting factor, the dry bulb temperature weightings generated for Vancouver, Victoria, and Abbotsford for the $TMY_{STATION}$ weather files are averaged. The climate zone-based weighting factors are then integrated into the Sandia method and the TMY_{CZ} weather files are generated using the historical weather data for each weather station. A total of 18 TMY_{CZ} weather files are generated using the climate zone-based weighting factors corresponding to the location's climate zone.

3.3.3 Building Simulation

The $TMY_{STATION}$ and TMY_{CZ} weather files generated in section 3.3.2 are simulated in their corresponding climate zone energy model, along with the CWEC weather files. The building energy demands are obtained from the energy simulations and are compared to evaluate the suitability of the $TMY_{STATION}$ and TMY_{CZ} weather files. Table 3.5 shows the total number of simulations run for each climate zone.

Table 3.5: Number of years simulated for each climate zone

Climate Zone	Abbreviation	Weather Files				Total
		$TMY_{STATION}$	TMY_{CZ}	CWEC	Long-Term	
4	CZ4	3	3	3	60	**69**
5	CZ5	3	3	3	58	**67**
6	CZ6	3	3	3	60	**69**
7A	CZ7A	3	3	3	60	**69**
7B	CZ7B	3	3	3	53	**62**
8	CZ8	3	3	3	39	**48**

The resulting monthly heating and cooling energy demands obtained from the simulations with the $TMY_{STATION}$, TMY_{CZ} and the CWEC weather files are compared to the long-term average heating and cooling demands.

3.3.4 Performance Evaluation

The ASHRAE Guideline 14 [20] requires the use of normalized mean bias error (NMBE) and the coefficient of variation of the root mean square error (CV(RMSE)) to evaluate energy model performance. The NMBE measures the amount of bias in the regression model, which indicates how closely the annual energy demand generated with the TMY weather files corresponds to the annual energy demand generated with the long-term weather data [20]. However, offsetting errors can influence the NMBE [20], therefore the CV(RMSE) is also used to determine the amount of variance in the regression model between the energy demand generated with the TMY weather files and the historical long-term data, indicating how well the TMY weather files and the long-term weather data align with each other [20]. The ASHRAE guideline requires the monthly NMBE values to be within ±5% and the CV(RMSE) values to be within ±15% for the energy model to be considered acceptable [20].

For this study, NMBE and CV(RMSE) are used to evaluate the performance of the proposed TMY weather files and the CWEC weather files. These metrics were selected since acceptable ranges for results verification are provided for the NMBE and CV(RMSE) by ASHRAE Guideline 14. Additionally, Reddy and Henze [21] recommend the use of the CV(RMSE) over the other popular metric, RMSE since it is a normalized value, therefore simplifying the comparison of the results.

The NMBE and the CV(RMSE) have been selected to evaluate the applicability of the $TMY_{STATION}$ and TMY_{CZ} weather files. The performance indicators investigate the amount of variance and bias in the building energy demand obtained from the energy simulations with the various weather files. The NMBE is calculated for the monthly energy demand using equation (3.2). A lower NMBE indicates a better fit to the long-term data.

$$NMBE = \frac{\sum_{i=1}^{N}(TMY_i - LTA_i)}{(n-1)*\mu} * 100 \tag{3.2}$$

Where:

TMY is the monthly energy demand results simulated using the typical meteorological year weather file.

LTA is the average monthly energy demand simulated using long-term weather data.

n is the total number of months in a year.

μ is the average monthly energy demand for the typical meteorological year weather file.

The CV(RMSE) is calculated for the monthly energy demand using equation (3.3). Additionally, a lower CV(RMSE) indicates a better fit to the long-term data.

$$CV(RMSE) = \frac{\sqrt{\frac{\sum_{i=1}^{N}(TMY_i - LTA_i)^2}{n-1}}}{\mu} * 100 \tag{3.3}$$

Where:

TMY is the monthly energy demand results simulated using the typical meteorological year weather file.

LTA is the average monthly energy demand simulated using long-term weather data.

n is the total number of months in a year.

μ is the average monthly energy demand for the typical meteorological year weather file.

3.4 Results

The Results section is divided into two main sections. The monthly weighting factors obtained from the proposed methodology used to generate the $TMY_{STATION}$ weather files and the TMY_{CZ} weather files are presented in section 3.4.1. In section 3.4.2, the building energy demands obtained from the energy model simulations and the performance indicator values calculated based on the energy demands are shown for the $TMY_{STATION}$, TMY_{CZ}, and CWEC weather files for each location.

3.4.1 Monthly Weighting Factors

The monthly weighting factors obtained from the machine learning model in Stage 3 of the proposed methodology are presented in this section. These monthly weighting factors will then be used to generate the $TMY_{STATION}$ and TMY_{CZ} weather files using the Sandia method.

3.4.1.1 $TMY_{STATION}$

The monthly weighting factors for each weather parameter were obtained using the proposed methodology for each weather station, as presented in Figure 3.4, and are organized by climate zone. The location-based weighting factors generated for each location can be found in Appendix B, Table B 1. All weather parameters exhibit variation, however distinct trends emerged when comparing the different climate zones.

Dry bulb temperature (DBT) weighting factors exhibited the most variation compared to the other weather parameters. Moreover, DBT received the largest weight across all months and locations, indicating its significant influence on energy demand. In most cases, the DBT weighting factor obtained using the proposed methodology was found to exceed the CWEC dry bulb temperature weight of 40%. There was a total of three instances where the DBT weighting factor was either less than or equal to 40%. This occurred for Vancouver in the month of October where DBT received a weighting of 30%, and for Iqaluit in June and August where the DBT weightings were 39% and 40%, respectively.

In Canada, winter extends from December to February, while summer spans from June to August. The remaining months, which occur during spring and autumn, serve as transitional periods, and are referred to as shoulder months. CZ4 has milder winters compared to the other climate zones and warm summers. In CZ4, DBT received higher weightings in the summer months and lower weightings during the winter months. The second most influential weather parameter for CZ4 was diffuse horizontal irradiance (DHI). The weight attributed to DHI was found to be inversely proportional to DBT, significantly increasing during the winter months. The weighting for direct normal irradiance (DNI) increased during the shoulder seasons compared to the other months. Wind speed (WS) and relative humidity (RH) received the lowest weightings, with a slight increase during the winter months. Additionally, there was some variation in the weighting factors between the three cities. Abbotsford received a slightly greater weighting for DBT across all the months compared to Victoria and Vancouver. Furthermore, as mentioned previously, Vancouver received a significantly lower DBT weighting in October, with an increase in DNI and DHI.

In CZ5 and CZ6, the weather parameters showed minimal variation between months compared to the other climate zones. These climate zones exhibit cold winters and hot summers, with CZ6 having a colder winter than CZ5. CZ6 receives a slightly larger weight on DBT compared to CZ5. All the cities within CZ5 have similar weightings to each other. In the case of CZ6, Ottawa and Montreal receive similar weights while St. Johns had consistently different weights.

CZ7A, CZ7B, and CZ8 exhibited similar trends to each other with CZ8 showing the most prominent trends. The trends were opposite to CZ4. These climate zones have much colder winters, with CZ8 being the coldest, and have milder summers compared to the other Canadian climate zones. The DBT weighting increased during the winter months, with CZ8 receiving the largest weighting for DBT compared to all other climate zones. Additionally, CZ8 has the largest amount of variation between the DBT weighting in the summer and winter. The remaining weather parameters received lower weightings during the winter months compared to the summer months. Furthermore, the cities within CZ7A and CZ7B received similar weighting factors. In CZ8, all the cities received similar weighting factors during the winter months, however during the summer months there was significant variation with the weighting factors with Iqaluit receiving a much lower DBT weight during June and August compared to the other two cities. Additionally, Yellowknife showed variation in weights compared to the other two cities in April, May, July, and October. The distinct trends emerging between climate zones may indicate the need for customized weighting factors based on local climate conditions.

Figure 3.5 further highlights the trends between the cities. As the climates get colder, the weights tend to increase during the winter months. Moreover, there is more variation in weights among the winter months between climate zones, with the weights ranging from 41% to 94%, while during summer they range from 39% to 78%, however the 39% and 40% weightings appear to be an outlier. Additionally, the figure further emphasizes the difference in weighting factors for St. Johns compared to the other CZ6 cities. St. Johns may warrant consideration for a different climate zone. Additionally, in summer, the Iqaluit DBT weights significantly differ from the values obtained for the other cities within CZ8. These differences in weightings between cities potentially indicate a need for refined climate zone definitions.

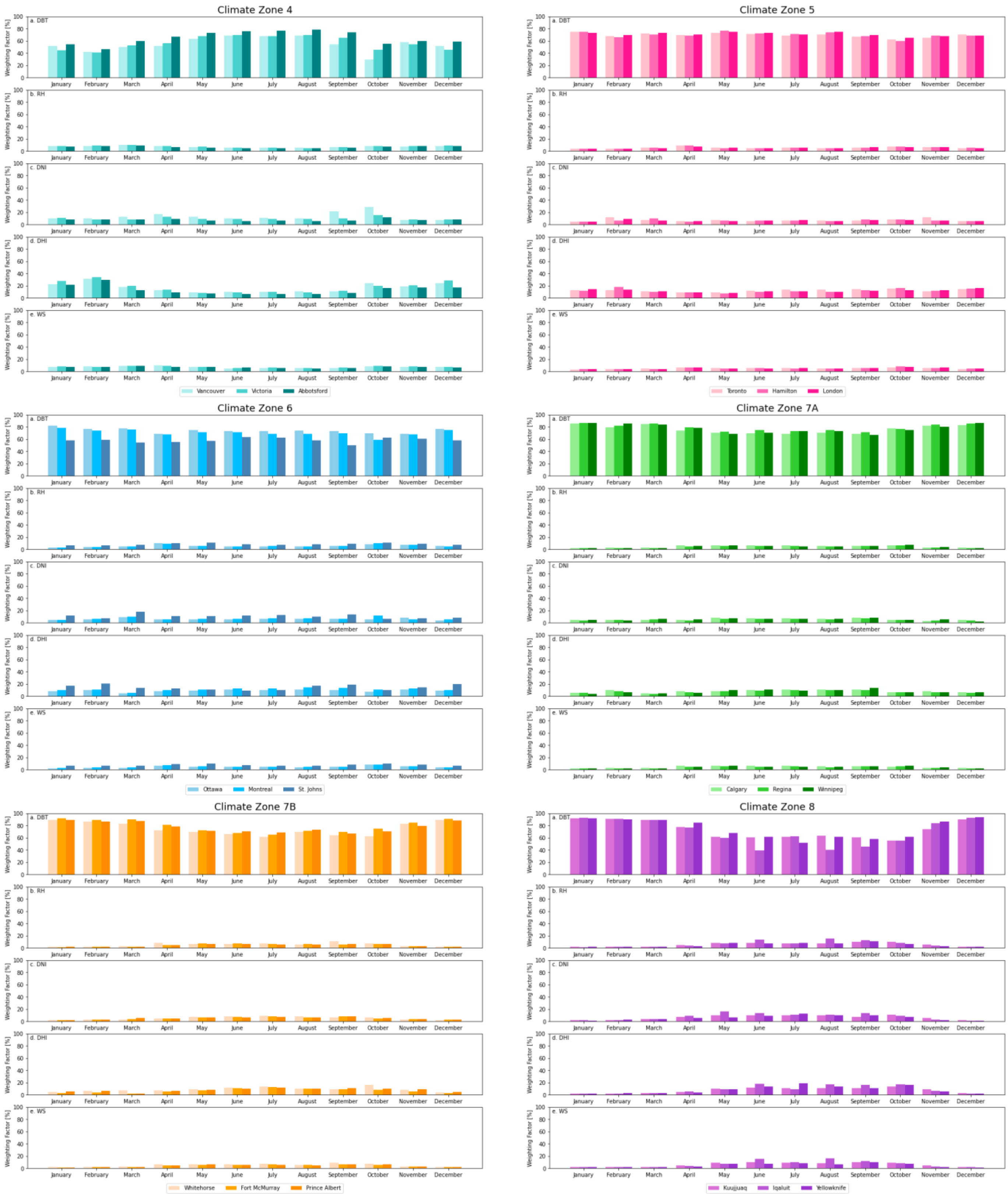

Figure 3.4: Weighting factors for each location sorted by climate zone

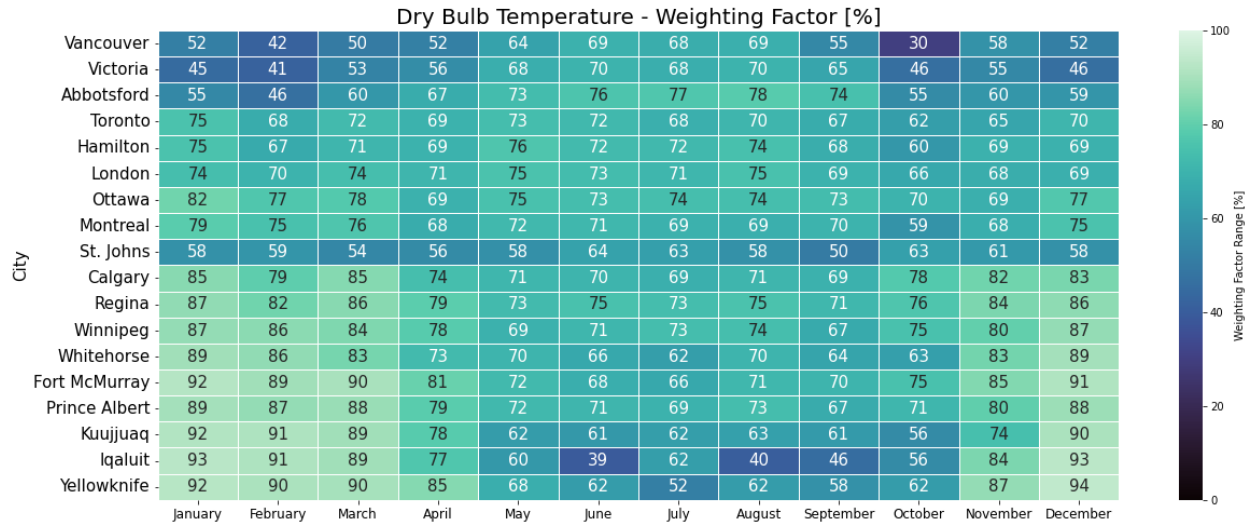

Figure 3.5: Weighting factors for dry bulb temperature for each city

3.4.1.2 TMY_{CZ}

As previously mentioned, a limitation of the framework, highlighted in [10], is the use of location-based weighting factors. As a result, a preliminary investigation is conducted in this study to evaluate the performance of standardized climate zone-based weighting factors. Figure 3.6 presents a heatmap of the averaged weighting factors for each climate zone organized by weather parameter. These average weighting factors will be referred to as climate zone-based weighting factors.

The climate zone-based weighting factors were used to generate the TMY_{CZ} weather file for each location using the location specific weather data in its respective climate zone. The climate zone-based weighting factors further highlight the trends shown between climate zones. CZ4 received the lowest DBT weighting during the winter months compared to the other climate zones, where CZ8 received the highest weight. Although CZ6 cities typically have colder winters then CZ5, CZ5 received a greater average DBT weighting than CZ6, this is most likely due to St. Johns, which was a potential outlier in CZ6. Both the CZ5 and CZ6 average weighting factors exhibited a lower variation in DBT weights compared to the other climate zones. DHI receives the largest weights during the summer in CZ8 and during the winter in CZ4. DNI experienced higher weightings in CZ8 during the summer, however CZ4 received the largest weighting during the shoulder seasons and experienced a significant increase in October. RH and WS typically received weightings under 10% except for in CZ8 where RH received a weight of 12% in September, and WS received a weight of 11% in June and September. Additionally, CZ8, receives both the minimum (2%) and maximum (12%) RH value. The climate zone-based weighting factors further highlight the monthly variation in weighting factors as well as the variation between the different climate zones.

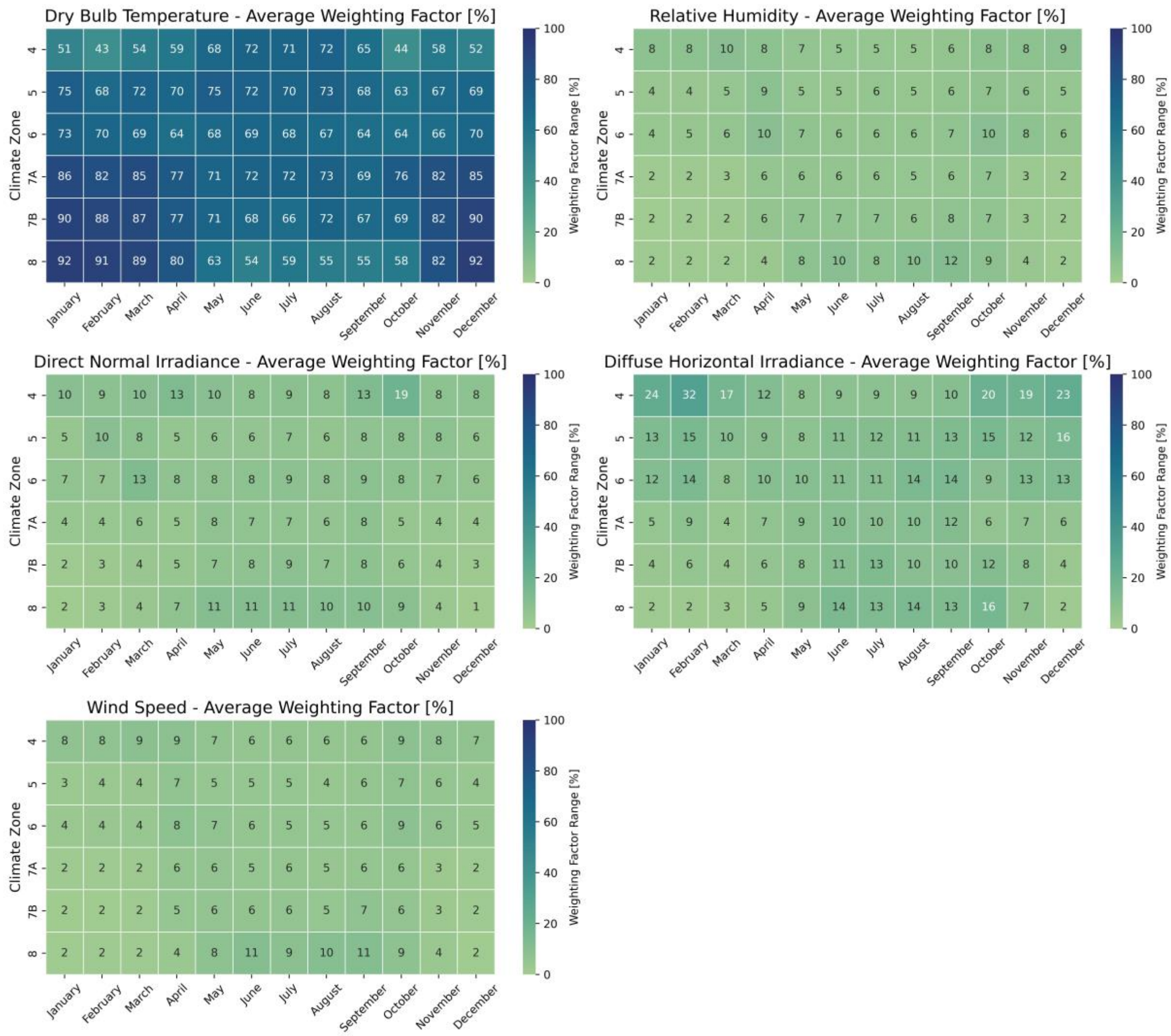

Figure 3.6: Average weighting factors per climate zone

3.4.2 **Energy Demand**

The monthly energy demands were obtained from the energy models simulated with the $TMY_{STATION}$, TMY_{CZ}, and CWEC weather files, along with the hourly long-term weather datasets for each location. The monthly energy demands were divided by the total building area and are displayed in Figure 3.7, Figure 3.8, and Figure 3.9. A difference in monthly energy demand between the $TMY_{STATION}$, TMY_{CZ} and CWEC weather files indicates months where different years were selected in the TMY weather file. Table B 2 in Appendix B displays the years selected for each month of the TMY weather files, with the years in bold indicating differences between the weather files. Seven of the proposed $TMY_{STATION}$ and TMY_{CZ} weather files are identical in composition, where the rest have one month where a different year was selected to represent the typical month.

The difference between each TMY weather file and the LTA varies across months and cities among the CWEC, $TMY_{STATION}$, and TMY_{CZ} weather files. The $TMY_{STATION}$ and TMY_{CZ} demands are very similar as the TMY weather file composition is very similar between the two. In most months across all cities, the proposed $TMY_{STATION}$ and TMY_{CZ} weather files appear to provide a better fit compared to the CWEC weather files. Moreover, discrepancies tend to be more pronounced in cases where the CWEC months demonstrate a poorer fit compared to the proposed $TMY_{STATION}$ and TMY_{CZ} weather files.

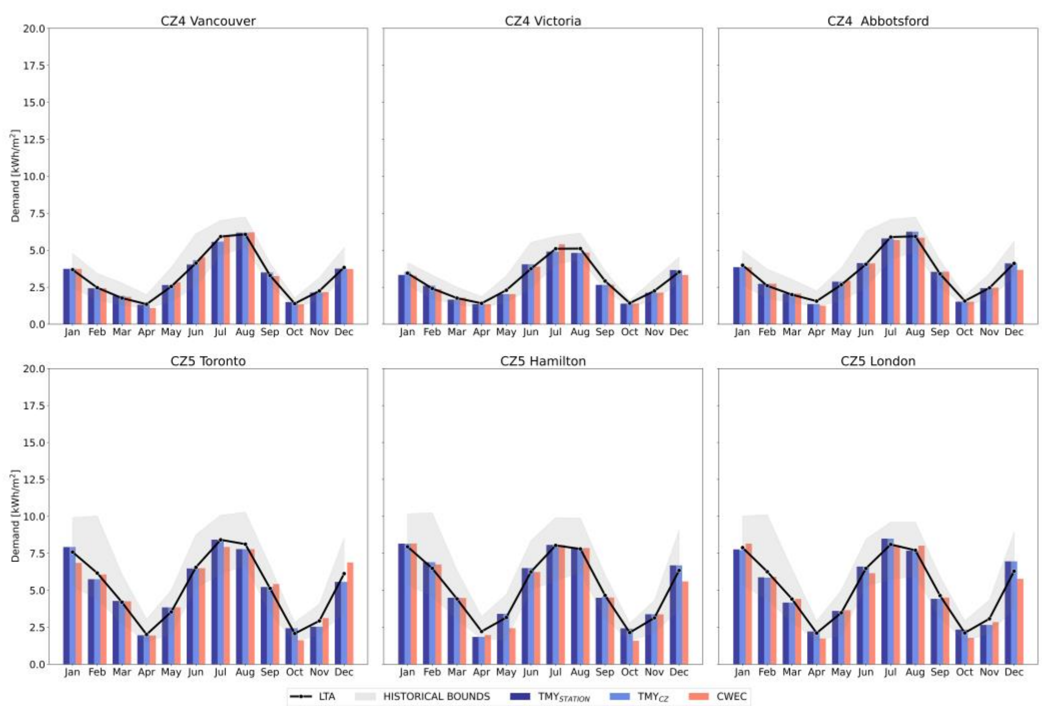

Figure 3.7: $TMY_{STATION}$, TMY_{CZ}, CWEC and LTA Energy Demand for CZ4 and CZ5

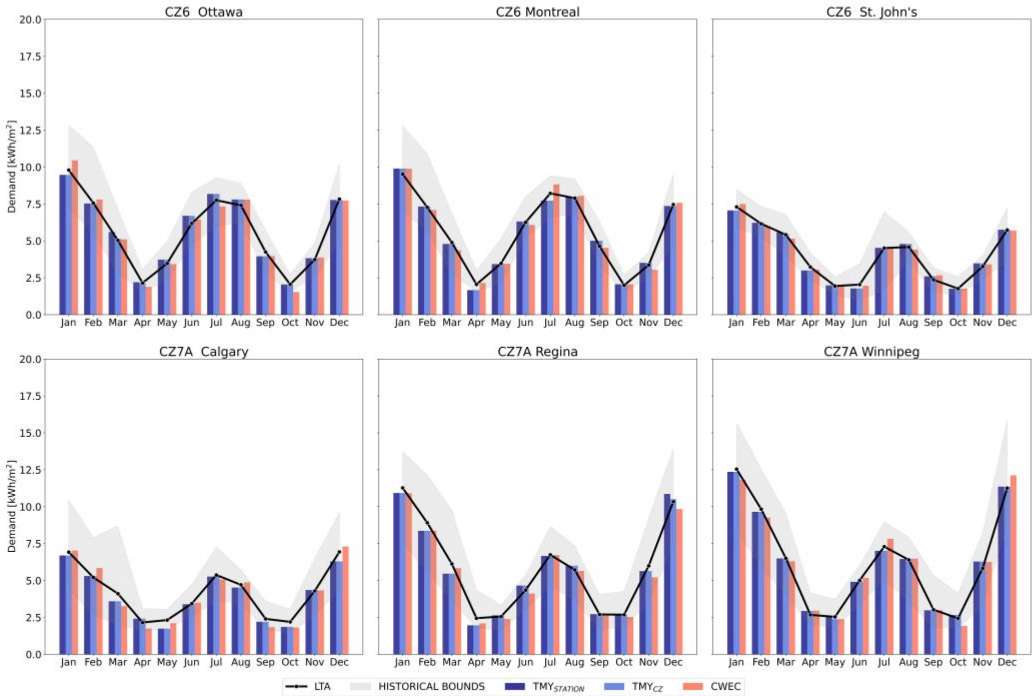

Figure 3.8: TMY$_{STATION}$, TMY$_{CZ}$, CWEC and LTA Energy Demand for CZ6 and CZ7A

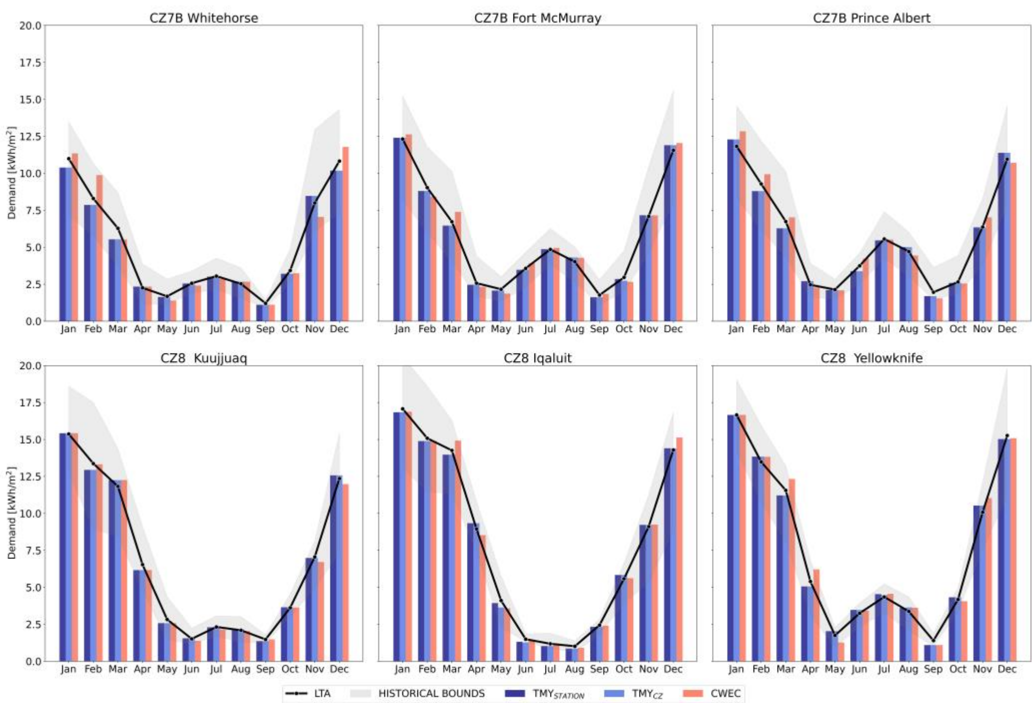

Figure 3.9: TMY$_{STATION}$, TMY$_{CZ}$, CWEC and LTA Energy Demand for CZ7B and CZ78

Table 3.6 displays the NMBE values and the CV(RMSE) values for CWEC, $TMY_{STATION}$, and TMY_{CZ} weather files in comparison to the historical long-term average weather data. The ASHRAE Guideline 14 requires the monthly energy demands NMBE to be within ±5% and the CV(RMSE) to be within ±15% for the energy model to be acceptable [20]. In most cities, the NMBE for all weather files remained within the acceptable range, except for Regina, simulated with the CWEC file, and Calgary, simulated with the $TMY_{STATION}$ and TMY_{CZ} weather files. Regina and Calgary exceeded the acceptable range by 0.7% and 0.5%, respectively, indicating a larger discrepancy between the long-term energy demand and the TMY weather file energy demand for these cities. Although still within the acceptable CV(RMSE) range, Calgary also received the highest CV(RMSE) value of 8.8%, indicating a higher variance in the results compared to the other cities. All the cities generated with the three different weather files had a CV(RMSE) within the acceptable range. Overall, the $TMY_{STATION}$ and TMY_{CZ} weather files had lower NMBE values and CV(RMSE) values compared to the CWEC weather files. The $TMY_{STATION}$ weather files had the most instances with the lowest NMBE and CV(RMSE) values, although the TMY_{CZ} values were either the same value or very close for many locations. These lower values represent a better performance with the $TMY_{STATION}$ and TMY_{CZ} weather files compared to the CWEC weather files, as they exhibited less bias and less variation in the results, indicating a better fit to the LTA weather data.

Table 3.6: Performance metrics

Climate Zone	City	NMBE			CV(RMSE)		
		CWEC	$TMY_{STATION}$	TMY_{CZ}	CWEC	$TMY_{STATION}$	TMY_{CZ}
		[-]	[-]	[-]	[-]	[-]	[-]
4	Vancouver	1.2%	0.0%	0.9%	5.5%	4.8%	4.9%
	Victoria	-3.0%	-2.6%	-1.9%	**6.7%**	7.0%	7.0%
	Abbotsford	-1.4%	1.2%	1.2%	6.4%	4.6%	4.6%
5	Toronto	**-1.1%**	-1.2%	-1.2%	8.0%	6.1%	6.1%
	Hamilton	-2.8%	2.7%	2.8%	7.6%	5.0%	4.9%
	London	-2.8%	0.4%	0.3%	**6.0%**	6.1%	6.2%
6	Ottawa	**0.2%**	2.5%	1.7%	6.2%	5.8%	5.0%

	Montreal	0.1%	0.0%	0.0%	5.5%	4.6%	4.6%
	St. John's	-0.4%	0.5%	0.3%	**4.2%**	4.5%	4.3%
	Calgary	**-3.0%**	**-5.5%**	**-5.5%**	10.7%	8.8%	8.8%
7A	Regina	**-5.7%**	-2.0%	-2.6%	6.9%	6.8%	6.4%
	Winnipeg	0.3%	0.5%	0.2%	7.6%	3.4%	3.5%
	Whitehorse	**0.9%**	-3.9%	-3.9%	13.4%	8.3%	8.3%
7B	Fort McMurray	1.4%	-0.4%	-0.6%	6.5%	3.3%	3.5%
	Prince Albert	2.8%	-0.6%	-0.6%	8.3%	5.9%	5.9%
	Kuujjuaq	-1.7%	-0.6%	-0.6%	3.9%	3.6%	3.6%
8	Iqaluit	**0.2%**	-0.7%	-1.1%	5.0%	2.8%	3.3%
	Yellowknife	2.9%	0.8%	0.8%	6.5%	3.9%	3.9%

3.5 Discussion

As revealed in the Results section, DBT consistently received a larger weight when compared to the other weather parameters indicating DBT has the greatest influence on building energy demand. The outdoor air temperature influences the amount of heat loss in the building. Mechanical heating and cooling are typically required to offset these losses contributing to the building's energy demand. The significance DBT has on the building's energy demand can be explained through heat transfer. Conduction can influence the heating and cooling demand of a building from heat loss or gains through the building envelope, such as the walls or roof. Fourier's law states the rate of heat transfer by conduction is directly proportional to the temperature differential [22]. The outdoor air temperature heavily influences the amount of heat loss that occurs through the building envelope. Additionally, heat loss or gain through convection occurs during infiltration. The amount of energy gained or lost from infiltration is partially influenced by the outdoor air temperature. Additionally, building ventilation requirements such as ASHRAE 62.1 [23], require outdoor air to be delivered during occupied hours, the outdoor air needs to be conditioned and the amount of sensible energy required depends on the differential between the outdoor and supply air temperatures. The flowrate of air entering a building and the temperature differential between the outdoor temperature and the target temperature influence the amount of sensible energy required to condition the air. The outdoor DBT significantly influences a building's energy consumption due to the energy required to condition and maintain space temperatures. Space conditioning, ventilation and infiltration can significantly impact energy demand which are all linked to the outdoor DBT, causing DBT to receive a larger weighting compared to the other parameters.

Table 3.7 provides the heating and cooling design day temperatures defined in NECB 2020 [15] for each weather station. Throughout the winter, the typical room temperature set point is 21°C. In Yellowknife, the temperature differential between outdoor and indoor conditions is -65°C, while in Vancouver it is -27°C. In cold climates, the large differential between the indoor and outdoor temperature, occurring during the winter months, may explain the significant increase in weight attributed to DBT compared to locations with milder winters. Similarly, a greater weighting is attributed to DBT during the summer months for climate zones which exhibit hotter summers. Furthermore, extreme cold climate zones like CZ7B and CZ8 also receive less hours of sunlight during the winter compared to the other climate zones, making temperature the primary driver of heating demand. Conversely, during summer, these zones receive more sunlight hours compared to the other climate zones and have lower summer temperatures, which slightly increases the solar weighting during these months. In CZ7B and CZ8, all other parameters increase during summer and decrease in winter. The differences in weighting factors and the trends observed between climate zones further highlight the need for weighting factors that account for local climate conditions.

Table 3.7: NECB 2020 Design Day temperature for each selected city [15]

Climate Zone	City	Design Day Temperatures Winter [°C]	Design Day Temperatures Summer [°C]
4	Vancouver	-9	28
	Victoria	-6	24
	Abbotsford	-10	29
5	Toronto	-20	31
	Hamilton	-19	31
	London	-20	30
6	Ottawa	-27	30
	Montreal	-26	30
	St. John's	-16	24
7A	Calgary	-32	28
	Regina	-36	31
	Winnipeg	-35	30
7B	Whitehorse	-43	25
	Fort McMurray	-40	28
	Prince Albert	-40	28
8	Kuujjuaq	-39	24
	Iqaluit	-41	17
	Yellowknife	-44	25

The Results section also discussed monthly variations in the weighting factors, these fluctuations were most prominent for DBT. Although the month-to-month variation is considered minimal, the weighting factors demonstrate significant seasonal variation. Figure 3.10 displays monthly boxplots of the dry bulb temperatures for each city. The seasonal fluctuations in weighting factors are likely attributed to the seasonal differences in the Canadian climate zones, consisting of four distinct seasons: fall (September to November), winter (December to February), spring (March to May), and summer (June to August). The seasonal variation in DBT, which can be seen in Figure 3.10, aligns with the variation in weighting factors. For instance, in the climates with the coldest winters such as CZ8, the DBT weighting is greatest during the winter, and gradually decreases during the summer, with the transition happening during the shoulder seasons. A similar trend is seen in CZ7A and CZ7B. CZ5 and CZ6 had minimal monthly and seasonal variation in weighting factors, possibly attributed to these climate zones experiencing hot summers and cold winters. Climate zones with the most pronounced seasonal variations, such as CZ4 and CZ8, experience a season that is more extreme than the other. The seasonal variation in the weighting factors for the various locations demonstrates the importance of monthly weighting factors. The current approach typically has one set of weighting factors that are applied for the whole year.

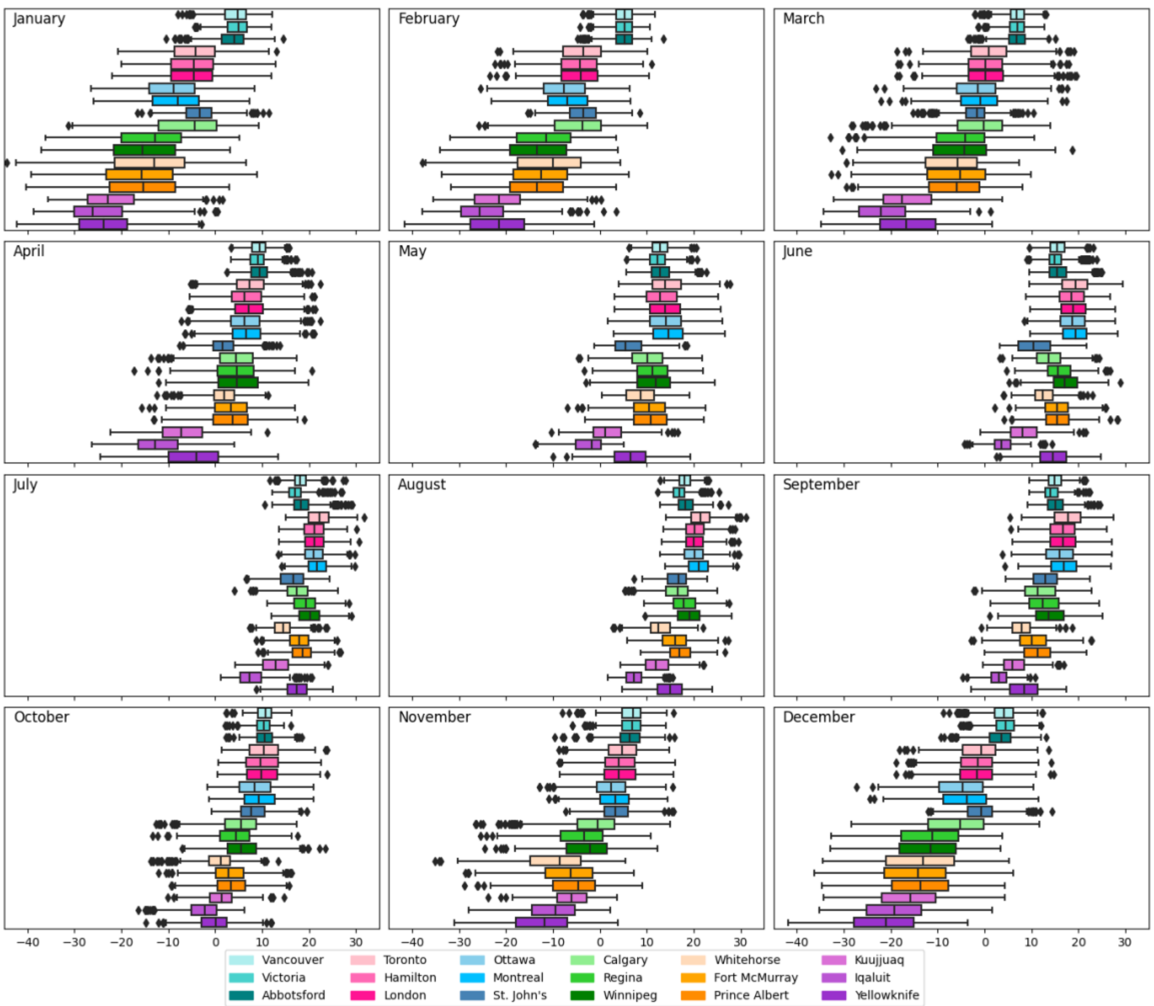

Figure 3.10: Boxplots of dry bulb temperature for each city

As previously discussed, distinct trends emerged when comparing the different climate zones. For instance, the colder climate zones receive a larger DBT weighting during the winter compared to the other climate zones. Similarly, the climate zones with the hotter summers receive a larger DBT weighting during the summer compared to the other climate zones, which is further supported by reviewing the DBTs for each location in Figure 3.10. These trends were further emphasized when the climate zone-based weighting factors were calculated. Notably, CZ5 was found to have greater DBT weights during the winter compared to CZ6, despite CZ6 typically having colder temperatures. This discrepancy is attributed to St. Johns, evident in Figure 3.10, having significantly different DBTs compared to the other CZ6 cities, which may influence the average weightings. The variations in weighting factors observed across different climate zones highlight the need for customized weighting factors which account for the different climate conditions.

Variations in weighting factors were also evident among cities within the same climate zone, potentially stemming from the diverse local weather conditions experienced by these cities. As shown in Figure 3.10, St. Johns and Iqaluit have different distributions in dry bulb temperature compared to the other cities within the same climate zone. These variations likely contribute to their difference in weighting factors in contrast to the other cities. Additionally, St. Johns is a coastal city whereas Ottawa and Montreal are located inland. As a result, the geographical difference may contribute to the disparities in weighting factors. Furthermore, the climate zones are defined based on heating degree days; therefore, the summer design day temperatures can vary significantly for each climate zone. Cities within the same climate zone which exhibit similar weighting factors were found to have similar design day temperatures, as is the case for cities within CZ5 and CZ7A. However, St. Johns and Iqaluit, have either different summer and/or winter design day temperatures. St. Johns consistently has different weighting factors compared to the other cities within CZ6 and has a significantly higher winter design day temperature and lower summer design day temperature compared to Montreal and Ottawa, as shown in Table 3.7. Iqaluit has different weightings during the summer months and has a significantly lower summer design day temperature compared to the other cities within CZ8. The variations in design day temperatures within a climate zone demonstrate the need for improved climate zone definitions. The significant disparities in weighting factors among cities within a climate zone highlight the limitations of employing an average weighting factor approach with the existing climate zone definitions.

The performance metrics indicated the $TMY_{STATION}$ and TMY_{CZ} weather files generally outperformed the CWEC weather files except for a few instances. Although CWEC exhibited a lower CV(RMSE) for Victoria, London, and St. Johns, the improvement was minor (within 0.3%), whereas the $TMY_{STATION}$ and TMY_{CZ} weather files showed up to a 5.1% improvement compared to the CWEC weather files. Overall, the $TMY_{STATION}$ and TMY_{CZ} weather files provided a more accurate representation of the monthly long-term average (LTA) energy demands, as supported by their lower CV(RMSE) values. The CWEC weather files had a total of five cities out of the 18 with a lower NMBE, indicating a better representation of the annual energy demand compared to the LTA for these locations. The CWEC NMBE values were up to 3.0% better compared to the $TMY_{STATION}$ and TMY_{CZ} weather files for these locations, whereas for the remaining 13 locations the $TMY_{STATION}$ and TMY_{CZ} weather files had up to a 3.7% and 3.1% improvement respectively. Additionally, for the locations where CWEC had a better NMBE, the $TMY_{STATION}$ and TMY_{CZ} weather files had better CV(RMSE) values for these locations which indicate less monthly error. Although CWEC had lower NMBE values for these locations, its higher CV(RMSE) values imply that the monthly variation in energy demands might have counteracted to achieve a closer approximation to the annual LTA energy demand. Furthermore, there are no instances where CWEC exhibited both a lower CV(RMSE) value and a lower NMBE value, whereas there are several instances when the $TMY_{STATION}$ and TMY_{CZ} weather files have lower values for both metrics. The $TMY_{STATION}$ weather files had the most instances where both the CV(RMSE) value and NMBE values were lowest compared to the other two files indicating a better performance. However, the performance of the TMY_{CZ} weather files differed from the $TMY_{STATION}$ weather files by at most by 0.9% for most locations. Lastly, all three weather files ($TMY_{STATION}$, TMY_{CZ}, CWEC) each had one location which exceeded the ASHRAE Guideline 14 acceptable range requirements by 0.5%, 0.5% and 0.7% respectively. In the case of the $TMY_{STATION}$ and TMY_{CZ} weather files, Calgary surpassed the threshold, possibly due to weather distribution, as evidenced by several outliers in the DBT boxplots for Calgary.

Although the $TMY_{STATION}$ and TMY_{CZ} weather files performed better than the CWEC weather files as indicated by the performance metrics, the amount of improvement across the cities within the same climate zone and in different climate zones was inconsistent. The CV(RMSE) values ranged from 2.8% to 8.8% and NMBE values ranged from 0% to 5.5% for the $TMY_{STATION}$ and TMY_{CZ} weather files. This variability in performance could be attributed to factors such as weather distribution, machine learning model accuracy, smoothing in the CWEC weather file, and the

Sandia method. Cities within the same climate zone may experience diverse local weather conditions, leading to variations in percentage improvements. These differences may be contributing to the variable performance in the weather files as well as the underperformance in a few. Moreover, the differences in dry bulb temperature within a climate zone further demonstrate the need for a refinement in the climate zone definitions. The machine learning model accuracy represents the machine learning algorithm's ability to predict the energy demand based on the weather inputs, however this does not necessarily represent the accuracy of the weighting factors. These accuracies can differ between months and cities due to the distribution in weather, which may contribute to the variation in performance between cities and climate zones. Furthermore, when CWEC weather files are created, six hours at the start and end of each month are smoothed using interpolation to remove step changes in the hourly data [24]. This smoothing may cause slight variations in the performance between the proposed weather files and the CWEC weather files. Lastly, the Sandia method is very sensitive to variations in the input datasets, such as a slight variation in rounding or weather station measurements can cause a different month to be selected. All these factors may attribute to the variations in performance between the cities. However, the overall improvement of the $TMY_{STATION}$ and TMY_{CZ} weather files indicates the need to consider customized weighting factors that account for local climate conditions.

The study found the methodology by Bigtashi et al. [10] to be effective in generating customized weighting factors for varying climates. The NMBE and CV(RMSE) results indicated the $TMY_{STATION}$ weather files better reflected the monthly long-term energy demand compared to CWEC for most of the cities. The improvement highlighted the importance of customized monthly weighting factors which account for local weather conditions. However, the study had a few limitations, such as only evaluating one building type and using a location-dependent methodology. Another study should be conducted evaluating the methodology on various building types. To reduce the time associated with generating location-dependent weighting factors, the study explored the use of a standardized set of climate zone-based weighting factors that were determined by taking the average of the weighting factors within each climate zone. The performance of the TMY_{CZ} weather files indicates the potential for a standardized set of climate zone-based weighing factors. Although the $TMY_{STATION}$ weather files performed slightly better than the TMY_{CZ} weather files, the convivence of the standardized climate zone-based weighting factors used in the TMY_{CZ} weather files would allow for the methodology proposed by Bigtashi et al. [10] to be more widely adopted. However, further refinement in climate zone definitions is necessary before these standardized weighting factor sets are created. Further analysis should be completed to consider a broader range of cities and diverse building types for an enhanced set of climate zone-based weighting factors.

3.6 Conclusion

Developing universal annual weighting factors based on expert judgement neglects variations in local climate conditions and seasonal weather fluctuations. While Bigtashi et al. [10] introduced a methodology that significantly improved TMY weather file performance when compared to the CWEC weather file, the study was limited to a single location and requires location-dependent weighting factors which can be time-consuming to generate for thousands of weather station locations. This study aims to address these limitations, with two primary objectives:

1. Assess the applicability of the machine learning methodology proposed by Bigtashi et al. [10] by applying it to various Canadian climates,
2. And to investigate the feasibility of employing a standardized set of climate zone-based weighting factors to reduce the computational time associated with generating location-based weighting factors.

The results demonstrate that the methodology by Bigtashi et al. [10] is versatile and can be applied to various climates. The $TMY_{STATION}$ weather files exhibited superior performance, demonstrating up to a 3.7% and 5.1% improvement in NMBE and CV(RMSE) values, respectively, compared to the CWEC weather files. Furthermore, the weather files generated using the standardized climate zone-based weighting factors (TMY_{CZ}) present a viable alternative, with the TMY_{CZ} NMBE and CV(RMSE) values being very similar to the $TMY_{STATION}$ weather files for most locations, with up to a 0.9% difference in performance between the two proposed weather files. The time saved by using the climate zone-based weighting factors to generate TMY weather files may be worth the marginal trade-off in performance between the $TMY_{STATION}$ and the TMY_{CZ} weather files. Moreover, the use of climate zone-based weighting factors can help reduce potential errors in situations where some locations may have low model accuracy due to weather distribution. While the TMY_{CZ} weather files provide a good representation of the long-term data, it is advisable to use the $TMY_{STATION}$ files for locations with weather patterns which significantly differ from other cities within the same climate zone. Such is the case for Iqaluit, where substantial disparities in weather distribution during the summer months are observed compared to other locations within the same climate zone.

Although both proposed TMY weather files performed better than the CWEC weather files, the study had a few limitations. The study focused only on a medium-sized office building and did not evaluate the methodology's performance in a warm tropical climate as it was limited to Canadian climate zones. Furthermore, the standardized climate zone-based weighting factors were generated based on a small sample size of three locations per climate zone. The limitations present opportunities for future work which may improve the proposed methodology by Bigtashi et al. [10] for cases of varying applications and climates. Furthermore, improving the definition of Canadian climate zones may improve the accuracy of the generated TMY_{CZ} weather files, as the climate zones are currently defined only based on heating degree days, and do not consider other variables such as cooling degree days. While climate zone-based weighting factors enhance the potential adoption of the proposed methodology, further research is imperative to redefine the climate zone definitions and expand the study to encompass diverse building types, sizes, and a broader spectrum of cities.

3.7 Acknowledgment

The authors acknowledge the support of the Natural Sciences and Engineering Research Council of Canada (NSERC) [RGPIN-2016-05711].

3.9 References

[1] City of Vancouver, "Energy Modelling Guidelines v2," City of Vancouver , Vancouver, 2018.

[2] City of Toronto, *Toronto Green Standard Version 4,* Toronto: City of Toronto, 2022.

[3] Government of Canada, "Engineering Climate Datasets," Government of Canada, 14 July 2023. [Online]. Available: https://climate.weather.gc.ca/prods_servs/engineering_e.html. [Accessed 24 September 2023].

[4] ASHRAE, "American Society of Heating, Refrigerating, and Air-Conditioning Engineers," ASHRAE, [Online]. Available: https://www.ashrae.org/technical-resources/bookstore/ashrae-international-weather-files-for-energy-calculations-2-0-iwec2. [Accessed 24 September 2023].

[5] Y. J. Huang, F. Su, D. Seo and M. Krarti, "Development of 3012 IWEC2Weather Files for International Locations (RP-1477)," *ASHRAE Transactions,* vol. 120, no. 1, pp. 340-355, 2014.

[6] The National Renewable Energy Laboratory, "Typical Meteorological Year (TMY)," NSRDB: National Solar Radiation Database, [Online]. Available: https://nsrdb.nrel.gov/data-sets/tmy. [Accessed 24 September 2023].

[7] W. Marion and K. Urban, "User's Manual for TMY2s," National Renewable Energy Laboratory , Golden, 1995.

[8] S. Wilcox and W. Marion, "Users Manual for TMY3 Data Sets," National Renewable Energy Laboratory, Golden, 2008.

[9] I. Hall, R. Prairie, H. Anderson and E. Boes, "Generation of Typical Meteorological Years for 26 SOLMET Stations," Sandia Laboratories, Albuquerque, 1978.

[10] A. Bigtashi, A. Papakyriakou and B. Lee, "Defining generation parameters with an adaptable data-driven approach to construct typical meteorological year weather files," *Energy & Buildings,* vol. 303, 2024.

[11] G. Georgiou, M. Eftekhari, P. Eames and M. Mourshed, "A study of the effect of weighting indices for the development of TMY used for building simulation," in *13th Conference of International Building Performance Simulation Association*, Chambery, 2013.

[12] T. Kalamees, K. Jylhä, H. Tietäväinen, J. Jokisalo, S. Ilomets, R. Hyvönen and S. Saku, "Development of weighting factors for climate variables for selecting the energy reference year according to the EN ISO 15927-4 standard," *Energy and Buildings,* vol. 47, pp. 53-60, 2012.

[13] T. Hong, W.-K. Chang and H.-W. Lin, "A fresh look at weather impact on peak electricity demand and energy use of buildings using 30-year actual weather data," *Applied Energy,* vol. 111, pp. 333-350, 2013.

[14] F. Meng, M. Li, J. Cao, J. Li, M. Xiong, X. Feng and G. Ren, "The effects of climate change on heating energy consumption of office buildings in different climate zones in China," *Theoretical and applied climatology,* vol. 133, pp. 521-530, 2018.

[15] Canadian Commission on Building and Fire Codes, National Energy Code of Canada for Buildings: 2020, National Research Council of Canada, 2022.

[16] ASHRAE, ASHRAE Standard 90.1-2022—Energy Standard for Sites and Buildings Except Low-Rise Residential Buildings, Peachtree Corners: ASHRAE, 2022.

[17] Canada Green Building Council, "Zero Carbon Building Design Standard Version 3," CaGBC, Ottawa, 2022.

[18] R. Morris, "Final Report - Updating CWEEDS Weather Files," Green Quill Climate Services Inc. , Auora, 2016.

[19] U.S. Department of Energy, "Commercial Reference Buildings," Office of Energy Efficiency & Renewable Energy, [Online]. Available:

https://www.energy.gov/eere/buildings/commercial-reference-buildings. [Accessed 9 October 2023].

[20] ASHRAE, ASHRAE Guideline 14-2023, Peachtree Corners: ASHRAE, 2023.

[21] T. A. Reddy and G. P. Henze, Applied Data Analysis and Modeling for Energy Engineers and Scientists, Springer, 2023.

[22] T. A. Reddy, J. F. Kreider, P. S. Curtiss and A. Rabl, Heating and Cooling of Buildings - Principles and Practice of Energy Efficient Design - Third Edition, Boca Raton: CRC Press, 2017.

[23] ASHRAE, ASHRAE Standard 62.1-2022: Ventilation and Acceptable Indoor Air Quality, Peachtree Corners: ASHRAE, 2022.

[24] Environment Canada, "Final Report - Updating CWEEDS Weather Files," Natural Resources Canada, Toronto, 2016.

Chapter 4 Conclusions and Future Work

The current approach to generating TMY weather files relies on weather parameters and weighting factors that are assigned based on expert judgement and often ignore seasonal variations and variations in climate and application. The proposed methodology provides a data-driven approach to define the weather parameters and weighting factors, accounting for seasonal variations by generating monthly weighting factors. The methodology was used to generate a proposed TMY weather file for Montreal using the building energy demand of a prototypical medium office building. The proposed TMY weather file outperformed the CWEC file by 16.04% with respect to total energy demand and the long-term average. These results demonstrate the importance of weighting factors that account for seasonal variations.

The methodology was applied to various locations across Canada to determine its applicability for different climates. The $TMY_{STATION}$ weather files demonstrated an improvement of representation in the LTA with up to a 3.7% and 5.1% improvement in NMBE and CV(RMSE) values, respectively, compared to the CWEC weather files for the majority of locations. The results confirm the proposed methodology is suitable for various climates.

Since the proposed methodology is location-dependent it can be time-consuming to generate customized weighting factors for each location. Therefore, the feasibility of using a standardized set of climate zone-based weighting factors was investigated. The TMY_{CZ} weather files also showed an improvement in the representation of the LTA for the majority of locations when compared to the CWEC weather files. The $TMY_{STATION}$ weather files showed a marginal improvement when compared to the TMY_{CZ} weather files with up to a 0.9% difference in performance. Although the location-based weather files performed slightly better for some locations, the convenience of using the standardized climate zone-based weighting factors to generate weather files may be worth the trade-off in terms of time efficiency.

Overall, the results demonstrate using a machine-learning methodology to generate TMY weather files can improve the accuracy in representing the long-term average when compared to the conventional approach to TMY weather generation. Therefore, using a machine learning methodology to generate TMY weather files proves to be applicable.

4.1 Thesis findings

1. Manuscript 1 demonstrates the improvement of TMY weather files in representing the LTA through integrating machine learning into the TMY weather file generation methodology.
2. Manuscript 2 improves the current approach to TMY weather file generation by demonstrating the applicability of the methodology to different climates and demonstrates the machine-learning methodology is suitable for different climate zones.
3. Manuscript 2 demonstrates the potential of using standardized climate zone-based weighting factors to help make the methodology more adaptable to industry applications.

4.2 Future Work

The TMY weather files generated with the proposed machine learning methodology demonstrated an improvement in performance across varying climates when compared to the conventional TMY weather file generation approach. To build upon the current research and further refine it, the following future work is recommended:

- Evaluate the proposed methodology's performance based on different applications.
- Evaluate how different energy simulation software may influence the results.
- Re-define Canadian climate zone definitions as they are currently only defined based on heating degree days. The new climate zone definitions should potentially consider cooling degree days, land use and the location such as coastal vs inland locations.
- Further, investigate the climate zone-based weighting factors by expanding the study to account for different building types and more locations.

References

[1] Government of Canada, "Green Building Principles," Government of Canada, 7 September 2023. [Online]. Available: https://natural-resources.canada.ca/energy-efficiency/green-buildings/green-building-principles/25301. [Accessed 14 February 2024].

[2] I. Hall, R. Prairie, H. Anderson and E. Boes, "Generation of Typical Meteorological Years for 26 SOLMET Stations," Sandia Laboratories, Albuquerque, 1978.

[3] National Climatic Center, "Typical Meteorological Year User's Manual,," 1981.

[4] H. Lund, "The design reference year users manual," Thermal Insulation Laboratory - Technical University of Denmark, Kongens Lyngby, 1995.

[5] R. Festa and C. F. Ratto, "Proposal of a numerical procedure to select Reference Years," *Solar Energy,* vol. 50, no. 1, pp. 9-17, 1993.

[6] S. Janjai and P. Deeyai, "Comparison of methods for generating typical meteorological year using meteorological data from a tropical environment," *Applied Energy ,* vol. 86, p. 528–537, 2009.

[7] Ontario Building Code, *O. Reg. 332/12, s.12.2.1.2.*

[8] British Columbia Building Code, *B.C. Reg. 264/12, s. 10.2.*

[9] City of Toronto, *Toronto Green Standard Version 4,* Toronto: City of Toronto, 2022.

[10] Environment Canada, "Final Report - Updating CWEEDS Weather Files," Natural Resources Canada, Toronto, 2016.

[11] The National Renewable Energy Laboratory, "Typical Meteorological Year (TMY)," NSRDB: National Solar Radiation Database, [Online]. Available: https://nsrdb.nrel.gov/data-sets/tmy. [Accessed 24 September 2023].

[12] W. Marion and K. Urban, "User's Manual for TMY2s," National Renewable Energy Laboratory , Golden, 1995.

[13] S. Wilcox and W. Marion, "Users Manual for TMY3 Data Sets," National Renewable Energy Laboratory, Golden, 2008.

[14] ASHRAE, "American Society of Heating, Refrigerating, and Air-Conditioning Engineers," ASHRAE, [Online]. Available: https://www.ashrae.org/technical-resources/bookstore/ashrae-international-weather-files-for-energy-calculations-2-0-iwec2. [Accessed 24 September 2023].

[15] Y. J. Huang, F. Su, D. Seo and M. Krarti, "Development of 3012 IWEC2Weather Files for International Locations (RP-1477)," *ASHRAE Transactions,* vol. 120, no. 1, pp. 340-355, 2014.

[16] M. Bhandari, S. Shrestha and J. New, "Evaluation of weather datasets for building energy simulation," *Energy and Buildings,* vol. 49, pp. 109-118, 2012.

[17] T. Hong, W.-K. Chang and H.-W. Lin, "A fresh look at weather impact on peak electricity demand and energy use of buildings using 30-year actual weather data," *Applied Energy,* vol. 111, pp. 333-350, 2013.

[18] Y. Cui, D. Yan, T. Hong, C. Xiao, X. Luo and Q. Zhang, "Comparison of typical year and multiyear building simulations using a 55-year actual weather data set from China," *Applied Energy,* vol. 195, pp. 890-904, 2017.

[19] K. Skeiker, "Comparison of methodologies for TMY generation using 10 years data for Damascus, Syria," *Energy Conversion and Management,* vol. 48, no. 7, pp. 2090-2102, 2007.

[20] World Bank Group, "Syrian Arab Republic," Climate Change Knowledge Portal, 2021. [Online]. Available: https://climateknowledgeportal.worldbank.org/country/syrian-arab-republic/climate-data-historical. [Accessed 2 March 2024].

[21] G. Georgiou, M. Eftekhari, P. Eames and M. Mourshed, "A study of the effect of weighting indices for the development of TMY used for building simulation," in *13th Conference of International Building Performance Simulation Association*, Chambery, 2013.

[22] H. D. Kambezidis, B. E. Psiloglou, D. G. Kaskaoutis, D. Karagiannis, K. Petrinoli, A. Gavriil and K. Kvadias, "Generation of typical meteorological years for 33 locations in Greece; Adaptation to the needs of various applications," *Theoretical and Applied Climatology,* vol. 141, no. 3-4, pp. 1313-1330, 2020.

[23] J. Sun, Z. Li, F. Xiao and J. Xiao, "Generation of typical meteorological year for integrated climate based daylight modeling and building energy simulation," *Renewable Energy,* vol. 160, pp. 721-729, 2020.

[24] T. Kalamees, K. Jylhä, H. Tietäväinen, J. Jokisalo, S. Ilomets, R. Hyvönen and S. Saku, "Development of weighting factors for climate variables for selecting the energy reference year according to the EN ISO 15927-4 standard," *Energy and Buildings,* vol. 47, pp. 53-60, 2012.

[25] F. Meng, M. Li, J. Cao, J. Li, M. Xiong, X. Feng and G. Ren, "The effects of climate change on heating energy consumption of office buildings in different climate zones in China," *Theoretical and applied climatology,* vol. 133, pp. 521-530, 2018.

[26] B. Qian, T. Yu, C. Zhang, P. Heiselberg, B. Lei and L. Yang, "A method of determining typical meteorological year for evaluating overheating performance of passive buildings," *Building Simulation,* vol. 16, no. 4, pp. 511-526, 2023.

[27] H. Li, J. Huang, Y. Hu, S. Wang, J. Liu and L. Yang, "A new TMY generation method based on the entropy-based TOPSIS theory for different climatic zones in China," *Energy,* vol. 231, p. 120723, 2021.

[28] A. L. S. Chan, "Generation of typical meteorological years using genetic algorithm for different energy systems," *Renewable Energy,* vol. 90, pp. 1-13, 2016.

[29] M. Kuhn and K. Johnson, Feature Engineering and Selection: A Practical Approach for Predictive Models, 1st ed., Boca Raton, Florida: Chapman and Hall/CRC, 2019.

[30] Canadian Commission on Building and Fire Codes, National Energy Code of Canada for Buildings: 2020, National Research Council of Canada, 2022.

[31] A. Bigtashi, A. Papakyriakou and B. Lee, "Defining generation parameters with an adaptable data-driven approach to construct typical meteorological year weather files," *Energy & Buildings,* vol. 303, 2024.

Appendix A – Chapter 2 Appendices

A.1 Proposed Methodology Flowchart

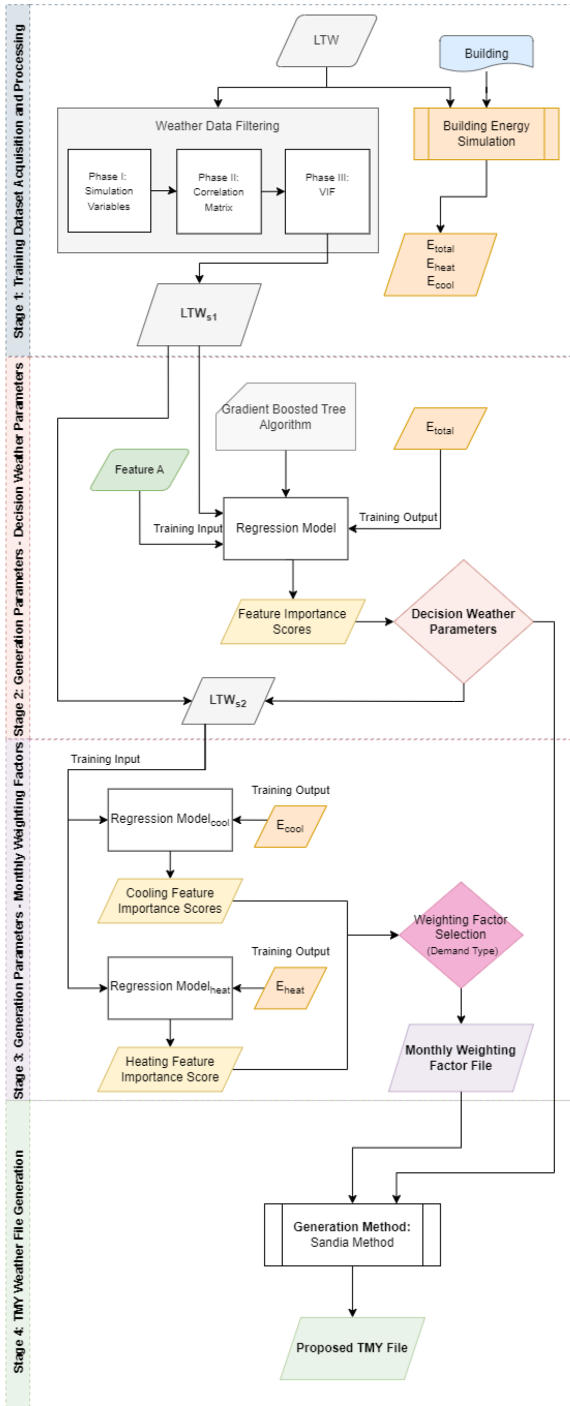

Figure A 1: Proposed methodology flowchart

A.2 Feature A

Feature A is used as a minimum threshold to evaluate the monthly feature importance scores for each weather parameter. The feature is created using a random number generator and integrated into the training input dataset. The purpose of integrating Feature A into the dataset is to introduce a variable which we are certain has no influence on the training output dataset, since it was independently generated and not considered in simulation. Therefore, features (weather parameters) with resulting feature importance scores equal or below that of Feature A are likely to have little influence on building energy demand.

However, although Feature A is independently generated, the probability that the randomly generated set of numbers demonstrates a significant correlation with the output dataset remains. In other words, there is always a possibility that two completely independent variables demonstrate a strong correlation. Therefore, to address this issue, the Stage 2 process is repeated for varying values of Feature A. Finally, the average feature importance scores from all repeated runs are used to evaluate and determine the relevant features (decision weather parameters).

A.3 Montreal Building Energy Demand

Table A 1: Monthly building energy demand comparison

Month	***Heating (kWh/m²)***			***Cooling (kWh/m²)***		
	CWEC	**Proposed TMY**	**LTA**	**CWEC**	**Proposed TMY**	**LTA**
January	9.90	9.90	9.50	0.00	0.00	0.00
February	7.08	7.09	7.28	0.01	0.01	0.01
March	4.27	5.00	4.81	0.08	0.07	0.11
April	1.57	1.57	1.35	0.59	0.59	0.70
May	0.04	0.04	0.09	3.41	3.41	3.39
June	0.00	0.00	0.00	6.08	6.78	6.27
July	0.00	0.00	0.00	8.83	8.56	8.19
August	0.00	0.00	0.00	8.05	8.02	7.89
September	0.01	0.01	0.02	4.52	4.51	4.63
October	1.12	1.12	0.99	0.94	0.94	1.00
November	2.90	3.65	3.28	0.14	0.08	0.10
December	7.57	7.32	7.45	0.00	0.00	0.00

Appendix B – Chapter 3 Appendices

Table B 1 Location specific weighting factors

Climate Zone	City	Weather Parameter	Weighting Factor (%)											
			January	February	March	April	May	June	July	August	September	October	November	December
4	Vancouver	DBT	51.72	41.78	50.13	52.10	63.93	69.10	67.80	68.60	54.84	29.89	58.44	51.86
		RH	8.19	8.22	9.88	8.16	6.73	5.48	5.57	5.36	6.45	8.67	7.84	8.68
		DNI	10.04	10.19	12.48	16.79	13.05	10.19	11.08	9.80	21.49	29.20	7.85	7.85
		DHI	22.41	31.83	17.95	12.73	9.27	10.15	10.02	10.72	11.23	23.92	18.67	24.10
		WS	7.64	7.97	9.57	10.22	7.02	5.08	5.54	5.52	6.00	8.32	7.19	7.51
	Victoria	DBT	44.76	41.48	52.79	56.38	68.25	69.63	68.09	69.92	64.87	45.92	54.69	45.88
		RH	8.07	8.83	10.31	8.34	7.14	5.95	5.50	5.17	6.37	8.76	8.49	8.97
		DNI	11.00	8.19	8.42	12.76	9.00	8.91	9.66	9.60	10.49	15.86	8.20	8.65
		DHI	28.08	34.03	19.56	13.46	8.66	9.56	9.96	9.37	11.78	20.02	20.41	28.68
		WS	8.09	7.47	8.92	9.06	6.95	5.96	6.79	5.94	6.49	9.44	8.21	7.82
	Abbotsford	DBT	55.03	46.24	60.12	67.49	72.97	75.92	76.89	78.48	74.30	55.50	59.76	59.04
		RH	7.79	8.32	9.56	7.00	6.02	4.85	4.63	4.57	5.42	7.25	8.01	8.54
		DNI	8.41	8.48	8.22	8.90	6.68	5.78	6.50	5.68	6.68	12.17	7.55	8.27
		DHI	21.49	29.61	12.83	9.59	7.26	6.93	6.36	6.22	8.18	16.66	17.41	17.39
		WS	7.28	7.35	9.27	7.02	7.07	6.53	5.63	5.06	5.43	8.42	7.27	6.76
5	Toronto	DBT	75.18	67.56	72.49	69.47	73.18	71.81	68.43	70.47	67.31	62.21	65.06	70.34
		RH	4.00	3.99	5.51	8.99	5.27	4.60	5.91	5.20	5.44	7.33	6.43	5.11
		DNI	5.11	12.21	7.07	5.44	7.03	6.08	6.85	6.32	6.96	7.98	12.07	5.62
		DHI	12.59	12.62	10.60	9.38	9.11	12.12	13.56	13.70	14.33	15.83	10.67	14.80
		WS	3.13	3.61	4.33	6.72	5.41	5.39	5.25	4.31	5.96	6.65	5.77	4.13
	Hamilton	DBT	75.08	66.53	70.62	69.21	76.44	72.29	71.55	74.30	67.79	59.96	69.08	68.85
		RH	4.18	4.20	5.45	9.32	4.84	5.11	6.02	4.91	6.00	7.21	6.33	5.57
		DNI	4.79	6.91	9.88	5.11	6.58	6.48	6.94	6.10	8.23	8.47	6.77	5.32
		DHI	12.28	18.44	9.79	9.61	7.02	10.51	10.63	10.25	12.45	16.33	12.17	15.66
		WS	3.67	3.91	4.27	6.76	5.11	5.61	4.86	4.44	5.53	8.02	5.66	4.60
	London	DBT	73.75	69.65	73.70	70.53	75.35	72.93	71.00	74.69	69.31	65.69	67.89	68.71
		RH	3.90	3.60	4.77	7.90	5.31	4.74	5.43	5.03	6.26	6.83	6.42	4.93
		DNI	4.54	9.45	6.72	5.93	5.88	6.36	7.50	5.73	7.45	7.12	6.39	5.97
		DHI	14.37	13.64	10.63	8.89	8.41	10.87	10.83	10.19	11.79	13.17	13.18	16.04
		WS	3.45	3.66	4.18	6.75	5.05	5.10	5.23	4.37	5.19	7.19	6.11	4.35
6	Ottawa	DBT	82.22	76.89	77.61	69.03	74.96	73.40	73.71	73.82	72.95	69.68	68.75	77.22
		RH	3.16	4.14	5.08	9.81	5.50	5.22	5.19	4.98	5.54	8.78	7.24	5.42
		DNI	4.35	5.92	9.38	5.85	5.82	6.03	6.58	6.24	6.43	5.81	7.92	4.22
		DHI	7.90	9.95	4.72	8.53	8.94	10.61	9.81	10.94	10.44	7.42	10.67	9.38
		WS	2.37	3.10	3.19	6.78	4.79	4.75	4.70	4.02	4.64	8.31	5.42	3.76
	Montreal	DBT	78.89	74.64	76.22	67.57	71.67	71.14	68.82	68.58	69.65	58.68	68.01	75.02
		RH	3.19	4.26	5.03	9.06	5.92	5.12	5.51	5.23	5.66	9.79	7.71	4.89
		DNI	4.70	6.49	9.81	6.00	6.41	6.53	7.73	7.14	6.50	12.04	5.70	6.04
		DHI	10.50	11.15	5.24	9.89	10.75	12.40	12.95	14.45	13.66	10.80	12.53	10.10
		WS	2.71	3.46	3.70	7.48	5.26	4.81	4.99	4.60	4.53	8.69	6.04	3.95
	St. Johns	DBT	58.10	59.18	54.40	55.70	57.63	63.90	62.88	58.02	50.26	63.01	60.82	58.11
		RH	6.77	6.74	7.23	10.48	10.65	8.12	7.46	8.04	9.60	10.74	9.29	7.28
		DNI	11.49	7.29	18.48	11.44	10.94	11.72	12.49	9.92	13.43	6.27	7.49	8.47
		DHI	17.34	20.67	13.29	12.86	10.83	9.09	10.37	17.15	18.80	10.09	14.53	19.49

		WS	6.30	6.12	6.60	9.52	9.96	7.17	6.79	6.86	7.90	9.89	7.87	6.65
7A	Calgary	DBT	85.44	79.43	84.99	73.76	71.07	70.00	68.60	70.70	68.92	77.70	82.06	83.28
		RH	2.36	2.90	2.77	6.40	6.20	6.24	6.44	5.84	5.84	6.22	3.37	2.97
		DNI	4.36	4.94	5.08	4.61	8.06	7.06	7.83	6.88	8.38	4.97	3.44	4.76
		DHI	5.61	9.91	4.49	8.67	8.27	10.51	10.75	10.79	10.81	6.14	8.05	6.33
		WS	2.24	2.82	2.67	6.56	6.40	6.19	6.37	5.80	6.05	4.97	3.08	2.66
	Regina	DBT	86.59	82.38	85.54	79.22	72.62	74.85	73.20	75.13	71.14	76.48	83.62	85.97
		RH	2.01	2.31	2.38	5.04	6.00	5.33	5.48	4.83	5.92	7.00	3.00	2.25
		DNI	3.70	4.40	5.71	4.00	6.87	6.14	6.45	5.79	7.22	4.67	4.15	3.91
		DHI	5.84	8.72	4.23	6.66	8.62	9.02	9.65	10.10	10.40	6.42	6.46	5.83
		WS	1.87	2.19	2.14	5.08	5.90	4.66	5.23	4.14	5.32	5.43	2.77	2.05
	Winnipeg	DBT	86.77	85.67	84.12	78.44	68.98	71.08	73.42	73.56	66.70	74.72	80.44	86.89
		RH	2.07	1.88	2.39	5.88	6.45	6.03	5.20	4.96	5.66	7.68	3.56	2.15
		DNI	5.16	3.68	6.50	5.44	7.59	6.75	6.71	6.37	8.19	4.69	5.62	2.23
		DHI	4.00	6.91	4.58	5.22	10.36	11.03	9.61	10.43	13.45	6.54	6.95	6.78
		WS	2.00	1.87	2.41	5.02	6.61	5.11	5.05	4.69	5.99	6.37	3.42	1.95
7B	Whitehorse	DBT	89.01	86.49	83.19	72.65	69.92	66.01	62.12	70.07	64.46	62.61	82.67	89.44
		RH	2.07	2.02	3.04	8.01	6.89	6.89	7.66	6.05	10.76	7.12	2.90	2.30
		DNI	2.35	2.63	3.41	4.88	7.57	8.12	9.67	8.13	6.50	6.46	3.10	2.14
		DHI	4.53	6.77	7.65	7.64	8.80	12.04	13.42	10.13	8.83	16.58	8.45	3.82
		WS	2.04	2.08	2.72	6.83	6.82	6.93	7.14	5.62	9.46	7.24	2.89	2.30
	Fort McMurray	DBT	92.10	89.03	90.11	80.91	72.21	68.07	65.75	71.30	70.10	74.78	84.72	91.11
		RH	1.63	2.01	2.05	4.76	7.26	7.23	6.91	6.26	5.89	6.42	3.06	1.92
		DNI	1.87	3.12	3.83	4.47	6.98	7.51	8.39	6.72	8.67	4.81	3.79	2.57
		DHI	2.93	3.94	2.32	5.25	7.68	11.29	12.49	10.06	8.97	8.67	5.87	2.72
		WS	1.47	1.90	1.69	4.62	5.87	5.90	6.46	5.66	6.37	5.32	2.57	1.68
	Prince Albert	DBT	89.41	87.05	87.88	78.88	71.79	70.68	69.15	73.34	67.10	71.08	79.98	88.02
		RH	1.74	2.12	2.26	5.21	6.27	6.75	5.78	5.49	6.54	6.86	3.48	1.97
		DNI	2.03	2.68	5.42	4.73	6.99	6.99	7.79	6.23	8.75	5.86	4.05	3.27
		DHI	5.23	6.28	2.46	6.44	8.65	10.34	11.79	10.02	10.98	9.68	9.52	4.85
		WS	1.59	1.86	1.98	4.74	6.31	5.24	5.49	4.92	6.63	6.52	2.97	1.89
8	Kuujjuaq	DBT	91.55	90.82	89.00	78.18	61.97	60.52	62.13	63.15	61.21	55.84	74.15	90.25
		RH	2.04	2.24	2.21	4.50	8.19	8.33	7.70	7.74	10.48	10.27	5.78	2.46
		DNI	2.06	2.55	3.88	7.83	9.95	9.95	9.78	10.32	7.38	10.76	5.60	2.01
		DHI	2.45	2.31	2.90	5.10	10.47	11.42	11.35	10.61	10.97	13.71	9.31	3.14
		WS	1.89	2.08	2.01	4.39	9.41	9.77	9.05	8.18	9.95	9.42	5.15	2.14
	Iqaluit	DBT	92.88	90.84	88.89	76.87	60.27	39.16	62.28	40.46	45.54	55.73	83.96	93.07
		RH	1.70	2.37	2.07	3.96	7.40	13.43	7.61	15.25	12.75	8.76	3.50	2.01
		DNI	1.73	2.53	4.23	8.86	16.08	13.78	10.84	10.99	13.68	9.65	3.08	1.02
		DHI	1.96	2.01	2.65	6.05	8.84	17.99	9.23	16.95	15.93	17.44	6.14	1.90
		WS	1.73	2.25	2.16	4.26	7.41	15.64	10.04	16.34	12.10	8.41	3.32	2.00
	Yellowknife	DBT	92.23	90.39	89.65	85.00	68.23	61.96	51.75	61.73	58.07	61.79	86.81	94.06
		RH	2.21	2.04	1.89	3.05	8.53	7.83	8.66	7.90	11.29	6.94	2.69	1.75
		DNI	1.63	2.70	3.84	5.37	6.71	9.14	12.41	9.85	10.23	7.31	2.36	0.94
		DHI	2.14	2.90	2.84	3.73	8.76	13.85	18.89	13.77	10.62	16.19	5.66	1.71
		WS	1.79	1.97	1.77	2.85	7.76	7.23	8.28	6.75	9.79	7.77	2.47	1.54

Table B 2 Generated TMY weather file selected years

Climate Zone	City	File Type	January	February	March	April	May	June	July	August	September	October	November	December
		TMY$_{STATION}$	2011	2000	2011	2010	2006	**2000**	2002	2015	2006	2004	2004	2011
	Vancouver	TMY$_{CZ}$	2011	2000	2011	2010	2006	**2017**	2002	2015	2006	2004	2004	2011
		CWEC	2011	2000	**2006**	**2013**	**2017**	**2016**	**2016**	2015	**2002**	**2009**	**2017**	**2001**
		TMY$_{STATION}$	**2014**	2012	2010	2010	2003	2017	2002	1999	2001	2000	2004	2000
4	Victoria	TMY$_{CZ}$	**2011**	2012	2010	2010	2003	2017	2002	1999	2001	2000	2004	2000
		CWEC	**2016**	**2008**	**2006**	**2015**	2003	**2000**	**2006**	1999	2001	2000	**2017**	**2003**
		TMY$_{STATION}$	2016	2006	2014	2007	2009	2000	2002	2012	2006	2000	2004	2011
	Abbotsford	TMY$_{CZ}$	2016	2006	2014	2007	2009	2000	2002	2012	2006	2000	2004	2011
		CWEC	2016	2006	2014	**2003**	2009	**2014**	**2010**	**2002**	2006	2000	**2017**	**2006**
		TMY$_{STATION}$	2010	2009	2007	2009	2006	2017	2015	2013	1999	2011	2004	1999
	Toronto	TMY$_{CZ}$	2010	2009	2007	2009	2006	2017	2015	2013	1999	2011	2004	1999
		CWEC	**2007**	**2004**	**2002**	2009	2006	2017	**1998**	2013	**1998**	**2014**	**2017**	**2004**
		TMY$_{STATION}$	2010	2008	2007	2004	2006	2002	2015	2012	2009	2011	2017	2016
5	Hamilton	TMY$_{CZ}$	2010	2008	2007	2004	2006	2002	2015	2012	**2011**	2011	2017	2016
		CWEC	**2000**	**2011**	2007	**2005**	**2009**	**2017**	2015	2012	2009	**2015**	2017	**2014**
		TMY$_{STATION}$	**2000**	2009	2007	2009	2006	2017	2013	2012	2009	2011	1999	2004
	London	TMY$_{CZ}$	**2016**	2009	2007	2009	2006	2017	2013	2012	2009	2011	1999	2004
		CWEC	**2010**	2009	**2006**	**2001**	2006	**2011**	**2008**	**2011**	**2008**	**2010**	**2006**	**2003**
		TMY$_{STATION}$	2012	2006	**2007**	2009	2006	2008	2015	2010	2001	2000	2000	2016
	Ottawa	TMY$_{CZ}$	2012	2006	2006	2009	2006	2008	2015	2010	2001	2000	2000	2016
		CWEC	**2011**	**2001**	2006	**2011**	**2014**	**2011**	**1998**	2010	2001	**2004**	**2004**	2016
		TMY$_{STATION}$	1999	2006	1999	2001	2006	2011	2003	2010	2007	2000	2000	2016
6	Montreal	TMY$_{CZ}$	1999	2006	1999	2001	2006	2011	2003	2010	2007	2000	2000	2016
		CWEC	1999	**2009**	**2009**	**2013**	**2014**	**2017**	**2013**	**1998**	**2001**	2000	**2016**	**2003**
		TMY$_{STATION}$	2017	2017	**2001**	2016	2000	2014	1998	2000	2014	2016	1999	2011
	St. John's	TMY$_{CZ}$	2017	2017	**2012**	2016	2000	2014	1998	2000	2014	2016	1999	2011
		CWEC	**2015**	**2006**	**2011**	**2005**	2000	**2001**	1998	**2016**	**2017**	2016	**2013**	**2009**
7A	Calgary	TMY$_{STATION}$	2002	2006	2000	2000	2012	2003	2000	2009	2000	1999	2011	2015

		TMY$_{CZ}$	2002	2006	2000	2000	2012	2003	2000	2009	2000	1999	2011	2015
		CWEC	**2003**	**2017**	**2001**	**2012**	**2007**	**2013**	**2001**	**2008**	**2016**	**2013**	2011	**2007**
		TMY$_{STATION}$	2010	2000	2004	2012	2015	2005	2015	2015	2005	2008	1998	**2016**
	Regina	TMY$_{CZ}$	2010	2000	2004	2012	2015	2005	2015	2015	2005	2008	1998	**1998**
		CWEC	2010	2000	**2008**	**2017**	**2000**	**2001**	2015	**2008**	**2014**	**1999**	**2008**	**2001**
		TMY$_{STATION}$	2007	2005	1998	2000	**1999**	2017	2003	2015	2017	2004	2007	2016
	Winnipeg	TMY$_{CZ}$	2007	2005	1998	2000	**2005**	2017	2003	2015	2017	2004	2007	2016
		CWEC	**2002**	**2004**	**2017**	2000	**2000**	**2011**	**2015**	2015	2017	**2008**	**2006**	**2017**
		TMY$_{STATION}$	2013	2006	2008	2014	2011	2011	2007	2005	2009	2010	2013	2009
	Whitehorse	TMY$_{CZ}$	2013	2006	2008	2014	2011	2011	2007	2005	2009	2010	2013	2009
		CWEC	**2011**	**2008**	2008	2014	**2017**	**2010**	**2011**	2005	2009	**2017**	**2008**	**2007**
		TMY$_{STATION}$	2014	2006	2004	2016	2001	2011	2000	2003	1999	**2017**	2002	1998
7B	Fort McMurray	TMY$_{CZ}$	2014	2006	2004	2016	2001	2011	2000	2003	1999	**1999**	2002	1998
		CWEC	**2003**	**2005**	**2017**	**2001**	**2007**	**2008**	**2001**	2003	**2002**	**2008**	2002	**2007**
		TMY$_{STATION}$	2014	2005	2007	2001	2011	1998	2011	1999	2005	2013	2002	2016
	Prince Albert	TMY$_{CZ}$	2014	2005	2007	2001	2011	1998	2011	1999	2005	2013	2002	2016
		CWEC	**1999**	**2009**	**1998**	**2017**	2011	**2011**	**2003**	**2000**	**2016**	2013	**2007**	**2001**
		TMY$_{STATION}$	2013	2006	2009	2013	2016	2006	2011	2015	2017	2008	2016	2014
	Kuujjuaq	TMY$_{CZ}$	2013	2006	2009	2013	2016	2006	2011	2015	2017	2008	2016	2014
		CWEC	2013	**2012**	2009	2013	2016	**2016**	**2010**	**2013**	**2006**	2008	**2012**	**2011**
		TMY$_{STATION}$	2007	2012	2009	2007	**2005**	2009	2005	2009	2005	2014	2007	2006
8	Iqaluit	TMY$_{CZ}$	2007	2012	2009	2007	**2014**	2009	2005	2009	2005	2014	2007	2006
		CWEC	**2009**	2012	**2008**	**2008**	**2016**	**2006**	**2012**	**2017**	**2016**	**2017**	2007	**2005**
		TMY$_{STATION}$	2010	2005	2012	2009	2013	2010	2017	2010	2016	2006	2013	2017
	Yellowknife	TMY$_{CZ}$	2010	2005	2012	2009	2013	2010	**2008**	2010	2016	2006	2013	2017
		CWEC	2010	**2011**	**2013**	**2008**	**2008**	2010	2017	2010	2016	**2007**	**2011**	**2014**

Appendix C – Chapter 2 Table Revisions

Table C1: Revisions to Table 2.2: Typical meteorological year weather file weighting factors

Weather Parameter	Measurement Parameter	CWEC [1][15] IWEC [13]	[TMY][1][4]	TMY2 [11], TMY3 [12], IWEC2 [14]
Dry-bulb temperature	Mean daily	30%	8.3%	10%
	Minimum daily	5%	4.2%	5%
	Maximum daily	5%	4.2%	5%
Dew point temperature	Mean daily	5%	8.3%	10%
	Minimum daily	2.5%	4.2%	5%
	Maximum daily	2.5%	4.2%	5%
Wind Speed	Mean daily	5%	8.3%	5%
	Maximum daily	5%	8.3%	5%
Global horizontal irradiance	Total daily	40%	50%	25%
Direct normal irradiance	Total daily	-	-	25%

Table C2: Revisions to Table 2.11: Monthly weighting factors for generation of proposed TMY weather file

Weather Parameter	***Jan***	***Feb***	***Mar***	***Apr***	***May***	***Jun***	***Jul***	***Aug***	***Sept***	***Oct***	***Nov***	***Dec***
DBT	34.63%	30.54%	30.53%	23.63%	45.03%	24.57%	17.68%	17.99%	34.82%	43.32%	25.51%	34.75%
RH	12.95%	11.35%	10.30%	11.00%	10.82%	10.26%	11.09%	10.35%	11.05%	11.29%	13.80%	13.21%
DNI	9.78%	10.10%	9.51%	16.52%	10.27%	11.69%	13.44%	11.14%	13.07%	13.29%	11.21%	9.64%
DHI	29.66%	36.98%	40.09%	39.20%	24.28%	43.46%	46.05%	51.02%	31.48%	17.34%	36.38%	29.17%
WS	12.97%	11.03%	9.57%	9.64%	9.61%	10.01%	11.73%	9.50%	9.58%	14.76%	13.10%	13.23%