Assessment of Urban Microclimate and Its Impacts on Building, Community, and Urban Energy Performance

Dongxue Zhan

A Thesis

In the Department

of

Building Civil and Environmental Engineering

Presented in Partial Fulfillment of the

Requirements for the Degree of

Doctor of Philosophy (Building Engineering)

Concordia University

Montreal, Quebec, Canada

March 2025

CONCORDIA UNIVERSITY

SCHOOL OF GRADUATE STUDIES

This is to certify that the thesis prepared

By: Dongxue Zhan

Entitled: Urban Microclimate Impacts on Building, Community, and Urban Energy Performance

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

Doctor Of Philosophy (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. Rabin Raut

______________________________ External Examiner
Dr. Yiqun Pan

______________________________ External to Program
Dr. Otmane Ait Mohamed

______________________________ Examiner
Dr. Bruno Lee

______________________________ Examiner
Dr. Mohamed Ouf

______________________________ Thesis Supervisor
Dr. Liangzhu (Leon) Wang

______________________________ Thesis Co-Supervisor
Dr. Ibrahim Galal Hassan

Approved by ____
Dr. Po-Han Chen, Graduate Program Director

March 7, 2025 ____
Dr. Mourad Debbabi, Dean of Faculty

ABSTRACT

Assessment of Urban Microclimate and Its Impacts on Building, Community, and Urban Energy Performance

Dongxue Zhan, Ph.D

Concordia University, 2025

With global efforts aimed at reaching carbon neutrality by 2050, there is an increased focus on improving the energy efficiency of buildings. The interactions between constructions and their local microclimate significantly influence the built environment and building energy performance. This thesis examines the urban microclimate and its impact on building energy consumption from the individual building level to entire urban areas.

Building energy models (BEMs) are essential for understanding building energy consumption, forecasting building energy, and evaluating energy-saving measures. Meanwhile Urban Building Energy Model (UBEM) is an analytical tool for modeling buildings on city levels and evaluating scenarios for an energy-efficient built environment. However, building planners commonly overestimate cooling loads by relying on Typical Meteorological Year (TMY) data in BEM/UBEM simulations, neglecting local microclimate variations and the neighborhood effects of surrounding buildings. This research developed an integrated platform by coupling BEM/UBEM with an urban microclimate model, allowing local aerodynamic data to be exchanged between the two models at each time step.

Since these BEM/UBEM models usually come with a deal of computation cost and prior knowledge to work with. In recent years, Machine Learning (ML) techniques in specific terms have been proposed for predicting building energy consumption. A synthetic dataset from physics-based simulations can serve as a training and testing data source for the ML model during the design phase. Weather clustering techniques are implemented to enhance computational efficiency and feasibility avoiding the high computational costs of day-by-day simulations. By employing weather clustering to select representative days, the approach reduces database size for training ML-based building prediction models.

The study begins with a comprehensive review of the latest methods for incorporating urban microclimate data into urban building energy models, addressing both methodological approaches and practical issues. Subsequently, the research evaluates the effects of urban microclimate on building energy performance, considering both individual buildings and urban-scale contexts. To address the computational cost associated with BEM/UBEM, an ML-based hourly building energy prediction model was developed, leveraging weather clustering techniques. The conclusion summarizes the key contributions of this thesis and offers recommendations for future research directions.

ACKNOWLEDGMENT

Throughout my PhD journey, I have been incredibly fortunate to receive an abundance of help and support. Without the guidance and encouragement of so many remarkable individuals, I would not have made it to this point. First and foremost, I would like to express my deepest gratitude to my supervisors, Dr. Liangzhu (Leon) Wang and Dr. Ibrahim Galal Hassan. I am deeply grateful to Dr. Wang for granting me the chance to work on multiple projects and for his unwavering guidance throughout my PhD. I also appreciate the significant amount of time he spent encouraging me and sharing invaluable insights that influenced not only my research but also my personal growth. My heartfelt thanks also go to Dr. Hassan for his invaluable advice, patience, and the substantial time he dedicated to my research. Their expertise and commitment to my academic development have been instrumental in shaping the quality of this work.

Furthermore, I would like to extend my appreciation to the collaborators at Lawrence Berkeley National Laboratory (LBNL), Dr. Tianzhen Hong, Wanni Zhang, and Dr. Na Luo, who offered invaluable support and assistance at various stages of both the project and my Ph.D. study.

A special thanks to Dr. Danlin Hou, Bingyan Jia, Dr. Shujie Yan, and Dr. Jiwei Zou, with whom I worked closely during the pandemic, one of the most challenging periods of my PhD. Your understanding and companionship during those times meant so much to me. I was also fortunate to collaborate with and learn from Dr. Senwen Yang, Majd Moujahed, Maher Albettar, Saeed Rayegan, Omar Ahmed, Dingyang Geng, Dr. Geng Tian, Wei Zhou, Dr. Yurong Shi, Peng Liu, Huiheng Liu, and Bahador Ziaeemehr, for the enjoyable moments we shared.

I am thankful to my dear friends, Yimeng Xia, Xinyu Zhou, Xianyao Fu, and Chenhan Wang, for all the wonderful memories we created together. A very special thanks to my partner, Hong Li, for your unwavering support, understanding, and constant encouragement. I also want to express my love and appreciation to my cats, Roke, Pangpang, Anan, and Bear, who have been my steadfast companions and a constant source of warmth and joy throughout this journey.

Finally, my heartfelt gratitude goes to my mother, Caiyun Wang. Your love and support have been my foundation, and I am forever grateful.

CONTENTS

List of Figures

List of Tables

Chapter 1. Introduction

1.1 Statement of Problem

Today, around 55% of the global population resides in urban areas, and this rate is projected to reach as high as 70% by 2050 [1]. Among various energy-consuming sectors, buildings account for over 40% of annual energy consumption worldwide and approximately 55% of electricity consumption [2], [3]. Thus, accurate estimation of building energy consumption is critical to provide city planners and policymakers with exquisite information on energy use to establish energy-efficient cities during the design phase. To this end, Building Energy Model (BEM) tools are widely used to estimate building energy consumption and investigate the influence of input variables on the energy performance of buildings [4]. BEM can simulate complex building physics provide explicit energy-saving performance and become an indispensable tool to explore the potential energy-saving of buildings. It considers the impacts of various input parameters, including the internal parameters, e.g., occupants, appliances, heating, ventilation, and air conditioning (HVAC) systems, and the external parameters, e.g., ambient weather conditions. The accuracy of the simulation results is subject to the uncertainties and importance of these inputs data. All input parameters in BEMs should be selected carefully to obtain accurate simulation results. Meteorological data is one of the most important types of information that has a substantial impact on building energy performance. However, Typical Meteorological Year (TMY) data have been widely used in BEM studies to represent the ambient climate of the building area without considering the local microclimate, omitting complex interactions between buildings and the environment [5]. The urban microclimate is a small area around a building that has different atmospheric conditions than the surrounding area. Buildings in urban areas suffer from higher air temperatures due to the UHI effect as well as reduced wind flow as a result of surrounding structures that block airflow. As a result of this, as well as reduced sky exposure and shaded solar heat, neighboring buildings alter the radiation balance [6]. These factors collectively affect the thermal and energy performance of urban buildings. A lack of knowledge of the local microclimate will decrease the accuracy of building energy simulation results.

A promising solution to investigate the impact of local microclimate on building energy use is to couple urban microclimate simulation tools with BEM tools due to there is no distinct tool that can directly assess the urban microclimate impact on building energy use [7]. Urban microclimate simulation tools predict local ambient conditions regarding different urban configurations. However, the features and the thermal processes of buildings are usually simplified or neglected in these simulation tools. BEMs can provide detailed descriptions of the building and its systems using a dynamic model in the building energy performance analysis involving many input parameters. There is a gap in explicitly quantifying the impact of local UHI on building energy consumption. Besides, compared to other inputs, the impacts of urban microclimate on BEM results are not adequately understood, and the literature also often presents inconsistent conclusions.

Since the BEMs/UBEMs usually come with a deal of computational cost and prior knowledge to work with. ML as a subset of artificial intransigence provides the ability to learn from data using computer algorithms. ML techniques contribute to bridging this gap by learning existing data to predict new samples and lead to informed decisions. They discover the relation between various input features and output targets (e.g. energy performance) using given data. ML techniques offer the flexibility to incorporate vast amounts of dataset gathered from various sources for predicting new samples and leading to informed decisions. Data sources could be smart grids, sensor networks, and on-site measurements, among others. These methods limit the computation complexity of the algorithm to computational time.

1.2 Motivation and objectives

Having briefly described the context and problem statement; the motivation and objective of this report are now introduced. The primary motivation of the proposed research work is to comprehensively understand the influence of local microclimate on building energy performance within the broader context of UHI effects. By conducting global sensitivity analysis, we are aiming to quantify the importance of urban microclimate and to identify the key parameters for the model output variations, study how significance of climatic parameters in building energy performance, compared against other building positive parameters, like envelop thermal properties, occupancy, and HVAC systems. And employing cutting-edge simulation techniques of BEM/UBEM and Urban Climate Model (UCM) in a coupling strategy at multiple scales, individual building and urban scale. The aim is to elucidate the significance of ambient temperature compared to other passive design parameters in building energy simulation (BES). Through this integrated approach, the research seeks to demonstrate how UHI impacts building energy performance, offering valuable insights for enhancing urban sustainability and resilience in the face of climate change. Moreover, considering the extensive dataset required for hourly building energy prediction models using simulated data, weather clustering techniques offer a promising solution to select representative days, thereby avoiding the high computational costs associated with day-by-day simulations.

Therefore, the strategy for the research will seek the following target:

- Develop novel coupling strategies to evaluate the local microclimate and its impact on individual building energy performance
- Propose a new integration platform for city-scale assessment of the urban environment during extreme weather conditions.

- Pioneering uses weather clustering techniques to reduce the size of datasets for the development of ML-based building energy prediction models.
- Develop a machine learning model to predict building energy consumption on an hourly basis considering multiple weather parameters and building features.

1.3 Thesis organization

To delineate the research gaps addressed in this study concerning urban microclimate and its implications, the thesis is structured into six chapters as follows:

Chapter 2 provides an up-to-date review of existing studies on coupling strategies between urban microclimates and building energy models.

Chapter 3 investigates the spatiotemporal characterization of urban microclimates and their impact on building-level energy performance through the development of an integrated platform that couples urban microclimate models with building energy simulations.

Chapter 4 introduces an hourly energy prediction model incorporating weather clustering techniques. By identifying a subset of representative days from yearly weather data, the model reduces computational complexity by focusing computational efforts on a reduced set of days.

Chapter 5 develops an integration platform linking urban microclimate models with urban building energy models, enabling city-scale holistic assessments of urban microclimates. The use of JSON schema facilitates information exchange between the models.

Chapter 6 concludes the research by summarizing the key contributions and identifying overlooked factors in the study. It also outlines potential directions for future research to enhance and expand upon the current work.

Chapter 2. Literature review

This chapter presents a comprehensive literature review on building energy modeling approaches across various spatial scales, ranging from individual buildings to urban-scale frameworks. It begins by reviewing conventional building energy models (BEMs) and then discusses the evolution toward urban building energy Models (UBEMs), which simulate energy consumption across large building stocks. However, both BEMs and UBEMs often overlook the influence of localized environmental factors, prompting the need to integrate urban microclimate models (UCMs). These models capture key phenomena such as the urban heat island effect and wind flow, which significantly impact building performance.

The chapter further explores the coupling strategies developed to integrate BEM/UBEM and UMCMs, enabling a more accurate representation of the dynamic interactions between buildings and their surrounding microclimate. Various one-way and two-way coupling approaches are reviewed, along with their strengths, limitations, and applications in recent studies[1]. In addition, this chapter reviews the growing role of machine learning techniques in building energy modeling. These data-driven methods offer promising solutions for rapid energy prediction, especially when accounting for ambient conditions as weather inputs.

2.1 Introduction

Urbanization is a global phenomenon, with a significantly increasing percentage of the world's population residing in urban areas. However, this rapid urban expansion has brought about various

[1] This chapter has included the contribution of the author in multiple publications:
1. Dongxue Zhan, Nurettin Sezer, Liangzhu (Leon) Wang, Ibrahim Galal Hassan (2023). "Coupling of Urban Microclimate and Building Energy Simulations: Review of the Recent Literatures." 6th International Conference on Countermeasures to Urban Heat Islands (IC2UHI), RMIT University, Melbourne, Australia.
2. Nurettin Sezer, Hamad Yoonus, Dongxue Zhan, Liangzhu (Leon) Wang, Ibrahim Galal Hassan, Mohammad Azizur Rahman (2023). "Urban microclimate and building energy models: A review of the latest progress in coupling strategies." Renewable and Sustainable Energy Reviews. Volume 184, 2023, 113577

challenges, including the Urban Heat Island (UHI) effect. The UHI effect refers to the increased temperatures experienced in urban areas compared to their surrounding rural regions. The consequences of the UHI effect exacerbate thermal discomfort, environmental pollution, serious health problems, and mortalities [8]. Moreover, the UHI effect significantly influences building energy demand [9] as the UHI-induced elevated air temperatures result in higher cooling demands and lower cooling effectiveness [6], which should be considered in the evaluation of building energy performance. Building Energy Models (BEMs) are used to evaluate the effectiveness of energy-conservation measures during the design stage [4]. All input parameters in BEMs should be selected carefully to obtain accurate simulation results. Weather data is one of the most important types of information that has a substantial impact on building energy performance. However, the measured climate data collected in the outlying rural areas are widely used in most existing BEM development methodologies [5], which ignores the meteorological difference between urban and rural areas. According to recent studies, the estimated building energy use greatly differs depending on whether the UHI impact is taken into account [10].

UHI significantly impacts building energy consumption [11]. Previous studies demonstrated the UHI effect based on the temperature obtained by both observation and simulation [12], For instance, Zinzi et al. [13] analyzed the urban climate in Rome, Italy using air temperature and relative humidity data from five weather stations across the city between October 2014 and September 2017. The average increase in air temperature due to the UHI effect was found to be 0.7 °C - 1 °C. Due to the UHI effect, the cooling load was calculated to increase by 53% and 74%, and the heating load decreased by 18% and 21% for the office building and residential building, respectively. Santamouris [14] reported that the cooling load of a typical urban building is 13% higher than a similar building in rural areas. UHI resulted in an average cooling load increase of 23% and a heating load reduction of 19%, which corresponded to an 11% increase in the overall energy use of a typical building. Another research by Li et al. [12] found that UHI could lead to a substantial increase in cooling energy demand between 10% and 120% and a reduction in heating energy demand between 3% and 45%. In addition, the UHI effect caused an increase in the median cooling load and heating load by 19.0% and 18.7%, respectively. It was indicated that the UHI effect may vary across the city. In contrast to the urban fringe, it was stronger in the urban core.

The approach and the information used to assess the UHI impacts in earlier investigations varied significantly. An approach to include the UHI effect in building energy models was put forth by Palme et al. [15]. The cooling demand for residential buildings in the coastal cities of the Pacific Ocean in South America was foreseen by downscaling the urban weather data at the building level of urban morphology. After incorporating the UHI effect, building energy demand increased in a range of 15%-200%. The study criticized the validity of current estimation studies, which overlooked the UHI effect. A similar study by Li et al. [16] investigated how UCM affects building energy use for air conditioning. By revising perceived temperature, the UHI effect, temperature and humidity effect, and cumulative effect were considered in estimating the air conditioning energy consumption. Microclimate had an impact of 11.3% increase in summertime electricity use, which peaked in 2005 at 20.4%. It was noted that the increased cooling load further intensively the UHI effect.

Given the significance of the UHI effect in simulating building energy performance, it is crucial to thoroughly examine the approaches implemented to incorporate this effect into building energy simulations. To this end, this work aims to review the literature on UCM and BEM tools, as well as the coupling strategies employed to integrate these tools effectively.

2.2 Methodology

Based on a systematic review methodology, the current work identifies the scope of the study, the research question, the literature resources, and the overall quality assessment.

2.2.1 Scope of the study

Most previous building energy simulation studies overlooked the local variations in atmospheric variables and their interaction with the built environment. Therefore, a detailed systematic study is required to compile and analyze the latest developments of microclimate impacts on building energy performance. The present review aims to fill this literature gap and help the readers understand the critical parameters that play a significant role during microclimatic interactions with urban buildings and choose suitable methods to accurately predict building energy performance in a real urban microclimate. The climatic conditions can be classified based on the spatial resolution, such as macroclimate (100–10000 km), mesoclimate (1–100 km), and microclimate (1 mm–1 km) (Figure 2-1). Urban microclimate describes the local climate effects that differ from the surrounding rural areas in terms of wind direction and speed, surface temperature, air temperature, and relative humidity [17], [18]. The majority of previous building energy simulation studies overlooked the local variations in atmospheric variables and their interaction with the built environment. Therefore, it is necessary to compile and analyze the latest developments in microclimate impacts on building energy performance in a comprehensive systematic study. The present review aims to fill this literature gap and provide readers with an overview of critical parameters that play a significant role during the microclimatic interactions with urban buildings and how to select appropriate methods to accurately predict building energy performance in a real UCM. The effect of UHI has been reviewed in several studies, such as the UHI intensity estimation in urban areas[19], UHI's impact on building energy [20], [21], the UHI interaction with heat waves [22], the intra-urban relationship between surface geometry and UHI [23], effect of spatiotemporal factors on UHI intensity [24], UHI mitigation strategies and tools [25], UHI mitigation policies and technologies [26]. A review of urban microclimatic impacts on building energy performance is presented, considering individual, community, and urban scale buildings. Microclimate and building energy modeling are discussed in detail, as well as software and coupling strategies that can be used to achieve better results.

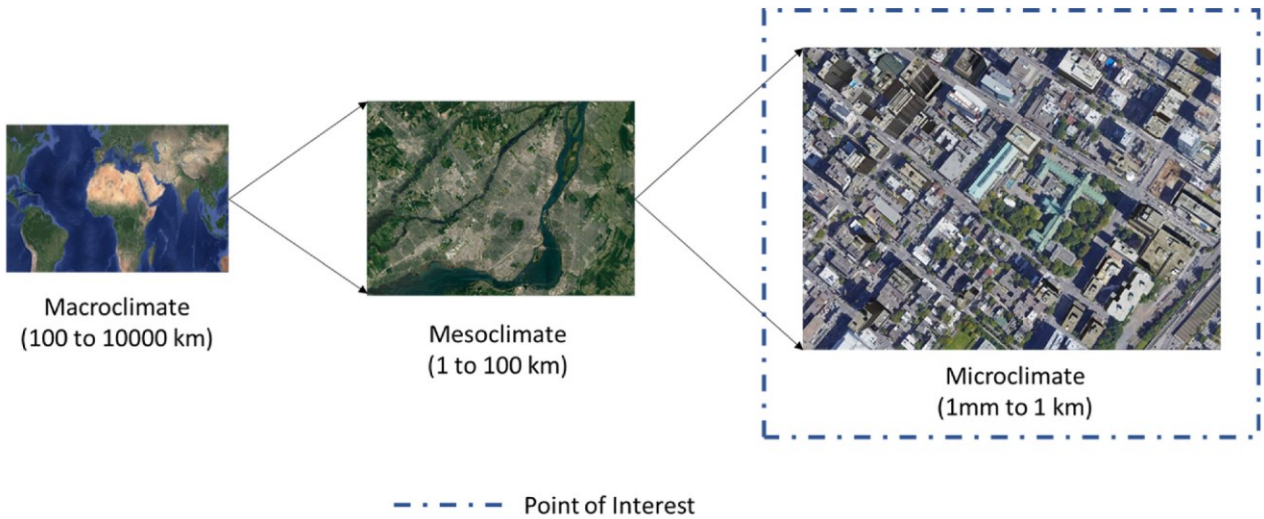


Figure 2-1. Identifying the point of interest of the present work. (Data source: Google Earth) [27].

2.2.2 Research questions

This study examines the following primary research questions: (1) how does microclimate affect the energy performance of a building? (2) How can the impact of microclimatic conditions on building energy consumption be effectively assessed? (3) How can UCM models be integrated into building energy models? (4) What are the available strategies for coupling UCM with BEM tools?

2.2.3 Collection and quality appraisal of the literature

Various steps were followed for the quality assessment. Research papers related to specific keywords are gathered through a primary literature search. Afterward, three levels of scrutiny are used to identify the most appropriate publications. Lastly, the resources are categorized and organized according to the sections of the paper. The papers are analyzed and summarized using the information that is most relevant or has the greatest impact on the topic.

Figure 2-2 presents the systematic screening process used to select relevant publications for the literature review using specific keywords. The initial search, based on a comprehensive set of keywords related to building energy simulation and urban climate, yielded 307 documents. Following that, three levels of scrutiny are used to sieve the resources and identify the most relevant publications and associated topics. The three levels of scrutiny are explained as follows. 1) Basic Scrutiny: Basic scrutiny is used to separate resources based on parameters like repetition, language, etc. The basic scrutiny is carried out by quickly skimming through the initially collected literature, 303 publications remained. 2) Intermediate Scrutiny: Within this scrutiny level, a more detailed analysis of the research papers based on the year of publication, type of publication (journal, conference, report, etc.), publisher, journal impact factor, cite score, etc. are used to evaluate the best literature material, reducing the dataset to 258 entries. 3) Strict Scrutiny: The last step of scrutiny contains a small pool of resources compared to the initial lot. The resources are scrutinized with ample time and removed mainly based on the breadth and depth of discussion, novelty, and adherence to the topic, resulting in a final selection of 243 high-quality studies. Finally, the resources are classified and organized into different sections of the paper.

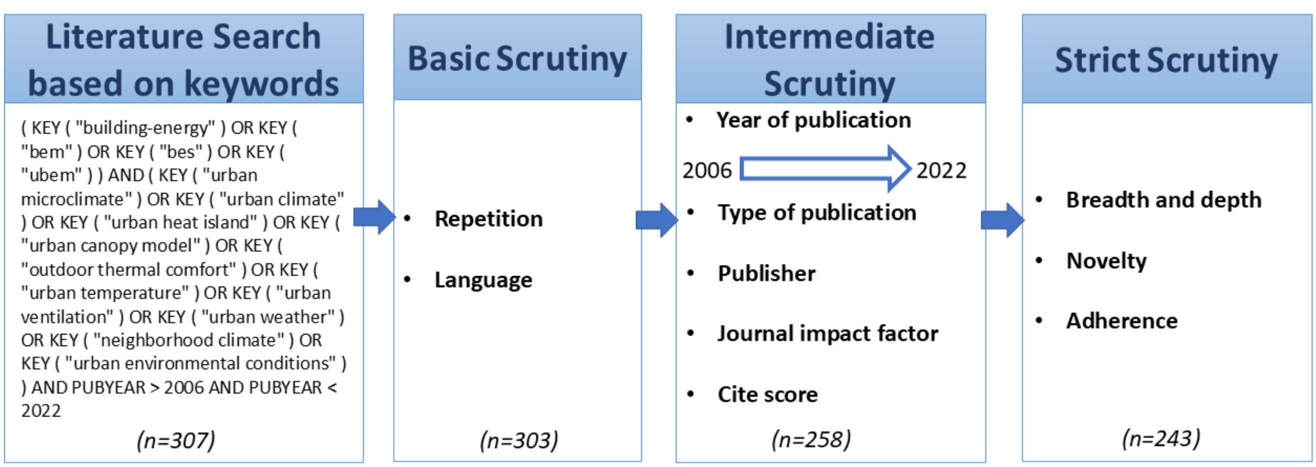


Figure 2-2. Steps of the quality assessment of resources.

The number of documents by year and journals on the literature search based on the is presented in Figure 2-3 and Figure 2-4, respectively. Figure 2-3 illustrates the annual number of documents related to the integration of energy models and urban microclimate from 2006 to 2022, based on Scopus data. There is a clear upward trend in publication volume, particularly from 2012 onwards, reflecting growing academic interest in this research area. The number of journal publications has consistently increased, reaching a peak in 2020 and remaining high through 2022. This trend highlights the expanding importance and relevance of coupled simulation methods in addressing urban energy climate challenges.

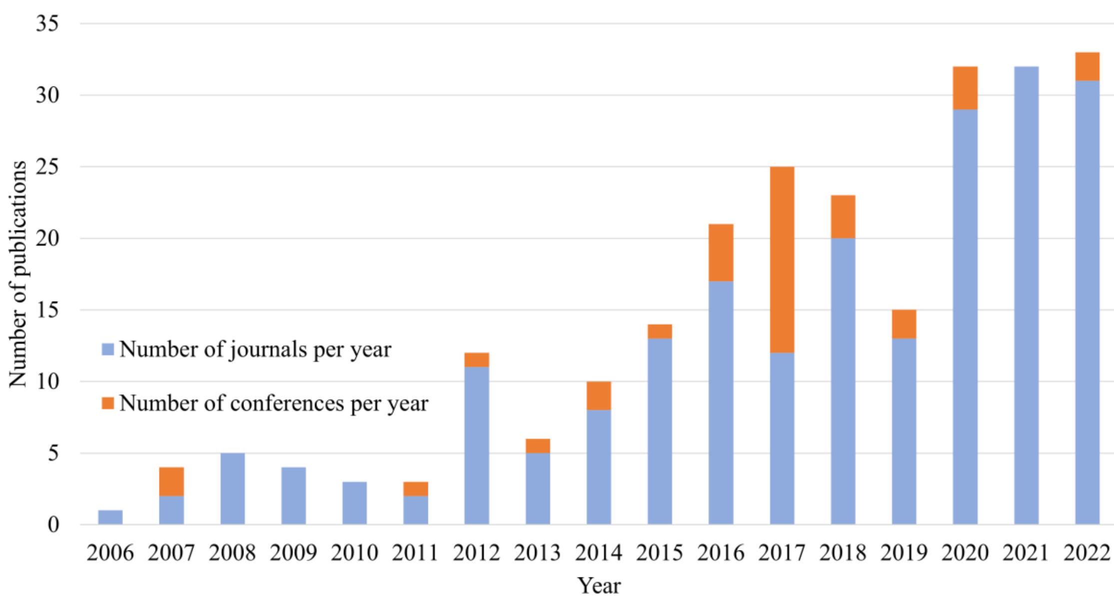


Figure 2-3. Number of documents by year (all data from Scopus).

Figure 2-4 shows the distribution of selected publications across different journals, based on data retrieved from Scopus. Among the journals, Energy and Buildings stands out as the most prominent source, publishing 50 relevant articles, followed by Building and Environment and Sustainable Cities and Society.

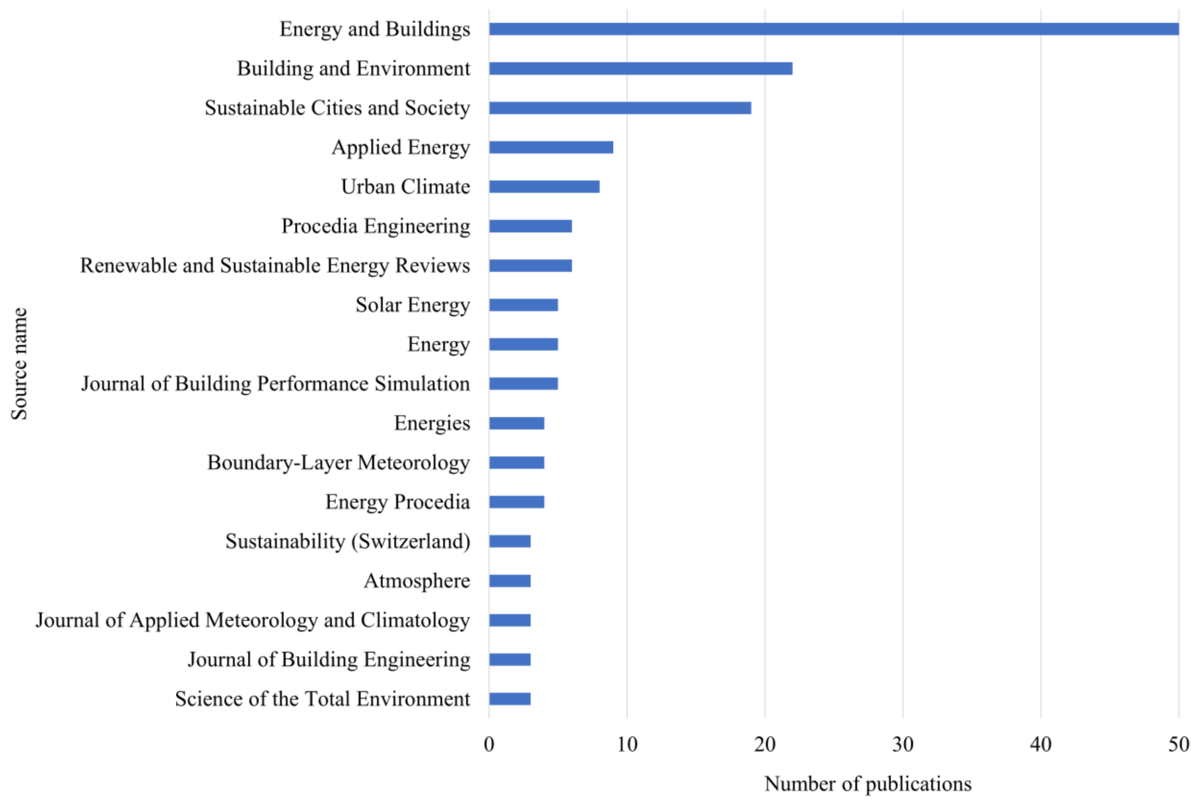


Figure 2-4. Number of publications by Journal (all data from Scopus).

2.3 Available simulation tools

Building energy modeling tools are developed for individual buildings as well as regionally. They were developed by universities and government agencies in the 20th and 21st centuries and have continually been improved with new versions released. Urban Climate Modeling (UCM) tools have newly developed models. In general, UCM software assists the users in modeling the surrounding urban environment based on reference weather station data and allows for conducting airflow analysis based on the Computational Fluid Dynamics (CFD) principles. Table 2-1 provides a list of various commonly used BES and UCM models with descriptions. All the listed UCM tools can be coupled with at least one BEM software.

Table 2-1. Overview of BES and UCM software.

SOFTWARE (Developer, year)	**FUNCTIONS**	**LINK TO PROGRAMS**
EnergyPlus (US Department of Energy, 2001)	• Three components - Simulation Manager, Heat and Mass Transfer Simulation Module, and Building Systems Simulation Module for integrated simulation. • Input - Building Description Data.	• Inbuilt link to TRNSYS and SPARK. • Functional Mock-up Units for co-simulation. Building Controls Virtual Test Bed (BCVTB) for coupling incompatible software.
ESP-r (University of Strathclyde, 1974)	• Separately model moisture, heat, electricity flow, and inter- and intra-zone airflow. • Handshaking of information takes place. • Input – Building geometry, atmospheric variables, and Convection Heat Transfer Coefficient (CHTC) values.	• Harmonizer for co-simulation with TRNSYS. • Other simulation tools can be linked using Python or other high-level programming languages.
TRNSYS (University of Wisconsin, 1975)	• Component-based workflow with defined parameters and input and output customized. • It is divided into two parts: the kernel or the simulator and the extensive model library of components with editable predefined models. • The models can be written in other programming languages and included in TRNSYS.	• It can be coupled with other simulation tools such as ENVI-met/Fluent using interfaces, e.g., Grasshopper or other script texts. • Internal link with Excel and MATLAB for other programs.
e-Quest (US Department of Energy, 2009)	• Hourly simulation of the building based on geometrical inputs. • Can perform a comparative analysis of various simulations from schematic to the final stage. • It also offers energy cost estimation, daylighting calculations, and intuitive energy-saving measures.	-
IES VE (Integrated Environmental Solutions, 1994)	• Modeling, thermal and load analysis, ventilation, lighting design, life cycle cost analysis, etc. • Uses the ApacheSim module for thermal simulation.	• MicroFlo is an internally integrated CFD module of IES VE.

SOFTWARE (Developer, year)	FUNCTIONS	LINK TO PROGRAMS
	• Allows sub-hourly simulation.	
CityBEM (Concordia University, 2019)	• The mass-energy conservation is solved for a whole building capsule with an indoor air cavity. • The building's overall thermal load is calculated.	• It is easily coupled with in-house CityFFD software.
Eco-Tect (Dr. Andrew Marsh, 2001)	• Hourly thermal comfort and monthly loads. • Daylighting, solar penetration, overshadowing, acoustic reflections and reverberations, project cost, and environmental impacts.	• WinAir plugin facilitates CFD analysis. • The ENVI-met solutions can be shared using EnergyPlus converters.
GreenBuilding Studio (AutoDesk, 2004)	• Optimization of energy performance, carbon use, and water use. • Compared with over 50 other solutions.	• It is linked to REVIT for a detailed building and environmental model.
ENVI-met (Dr. Michael Bruse, 1993)	• Solar and air pollutant dispersion analysis, building energy performance, green and blue area analysis, and outdoor thermal comfort level.	• BCVTB module or Meteonorm can be used to couple with EnergyPlus. • The TRNSYS outputs can be combined using Grasshopper.
Fluent (ANSYS, 2006)	• Fluid flow and transport phenomena: Incompressible, compressible, laminar, and turbulent fluid flow. • Steady and transient state analysis with natural, forced, and mixed convection. • Porous media, multi-reference flow, swirl, lumped parameter, and stream-wise periodic flow and heat transfer.	• Through BCVTB and MATLAB to link EnergyPlus. • MATLAB-Fluent coupling allows easy coupling with other software. • Direct Coupling of Fluent and SOLENE code.
CityFFD (Concordia University, 2019)	• Hourly Typical Meteorological Year (TMY) data is used to simulate local variations in atmospheric variables based on reference weather station values. • Higher-order forward and backward sweep and interpolation of the solver.	• It is easily coupled with CityBEM.

SOFTWARE (Developer, year)	FUNCTIONS	LINK TO PROGRAMS
UWG (Massachusetts Institute of Technology, 2012)	• Estimate hourly urban canopy temperature and humidity based on the weather station data input. • Input EPW weather file and XML file of urban geometry output the effects of UHI in EPW.	• As the output file is in EPW format, it is directly compatible with most BES software. • Usually used for normal chaining.
SOLENE-Microclimate (CERMA Laboratory, 1990s)	• It consists of three models: Thermal, CFD, Radiation • Quantify envelope materials and impacts of landscapes.	• Direct chaining with BuildSysPro. • Submodels for SOLENE are developed as a building thermal model to be coupled with Fluent.
OpenFoam (Henry Weller, 1989)	• Meshing, discretized operators, and physical models in the form of predefined libraries. • Code functionality to implement complex physical models.	• It can couple with EnergyPlus and similar software with the help of middleware.

2.4 Coupling strategies

There is no distinct tool that can directly assess the UCM impact on building energy use. Coupling UCM and building energy simulations can yield reliable predictions of microclimate impact. A coupling is the rapid exchange of information or variables between different computational engines that helps to solve the necessary equations urged by the solvers. This section intends to help the readers identify the methods, guidelines, and limitations of available coupling strategies. The software can be coupled in a variety of ways. The coupling strategies can also be classified based on the flow of variables, the capacity of the solvers, and the operation of timesteps. The research papers explicitly identify one-directional and two-directional coupling strategies and external coupling is most prevalent, so this review follows the same classification. This article discusses these two methods in detail. Afterward, the guidelines and limitations of these methods are discussed.

2.4.1 One directional coupling

One-directional coupling strategy is a method of sharing values for unknown variables from one software to another without feedback. Figure 1 summarizes the one-directional flow of information for the studies explained below.

One-directional coupling strategy is a method of sharing values for unknown variables from one software to another without feedback. Figure 2-5 summarizes the one-directional flow of information for the studies explained below. Liu et al. [28] implemented a one-directional coupling of CFD and BEM to investigate the impact of UCM on the energy performance of an academic building in the US. CFD software was used to generate local velocity and temperature variables based on velocity and temperature variables from the Energy Plus Weather (EPW) data. Microclimate variables were stored as EPW files with these local variables. The EPW file was entered into one of the physical models to solve the heat transfer variables on the BES platform. This is a one-way flow of information, and CFD and BES platforms are used separately to model microclimate and heat transfer.

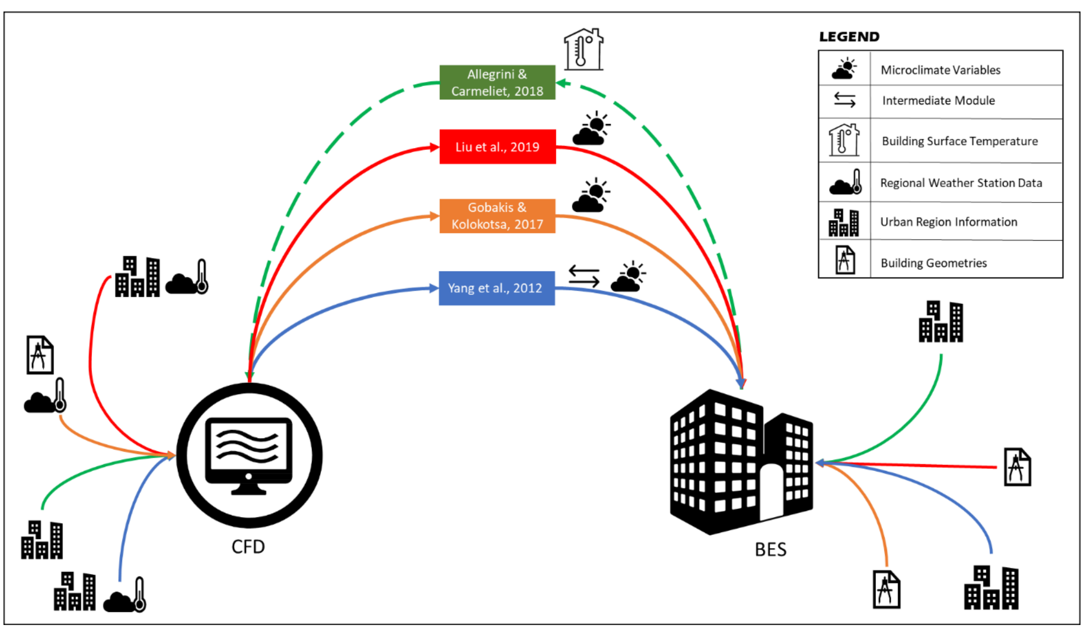


Figure 2-5. One-directional exchange in literature.

Furthermore, a module can be used to incorporate two independent platforms in addition to the aforementioned manual coupling strategy. Yang et al. [29] used a Building Controls Virtual Test Bed (BCVTB) for coupling ENVI-met with EnergyPlus. To satisfy the boundary condition of EnergyPlus, an 'E-E module' transfers hourly physical variables from ENVI-met. The building envelope is defined using compatible units for both models. Microclimate data with each unit is averaged and transferred to EnergyPlus. The outside boundary conditions remained the same for all linking units. For the initial time step, EnergyPlus sends unknown values such as air temperature to the E-E module. This module collects information and physical variables from the ENVI-met microclimate simulation to prepare heat transfer coefficients. These coefficients are shared with EnergyPlus for the simulation, and the iteration continues.

A case study in Switzerland investigated the impact of local microclimate on building energy consumption [30]. The study used CFD-BES software coupling with one-directional communication. Building surface temperature served as the CFD boundary condition, while no feedback was utilized from CFD to BES. The radiation model accounted for shortwave and longwave radiations within the BES software. Heat transfer employed an electric circuit equivalence and correlations from literature. Multiple simulations were conducted, varying wind speeds, directions, albedo values, and designs. The study aimed to understand microclimatic variations around the building compared to ambient temperature. Another study indirect coupling between BES and CFD to assess the impact of the urban environment on building energy performance and indoor environmental quality [31]. The thermal model of university buildings was developed in ESP-r software, and ENVI-met was used for microclimate simulation. Weather variables from ESP-r were updated and stored, while CHTC values were calculated using different methods. The coupling of ENVI-met and ESP-r was achieved through Python programming. To accommodate CHTC values, the simulation was divided into seven periods and dynamically run in ESP-r solver.

2.4.2 Two directional coupling

Two directional coupling strategies are the exchange of information between the software, contributing to the unknown values required for the simulation (Figure 2-6).

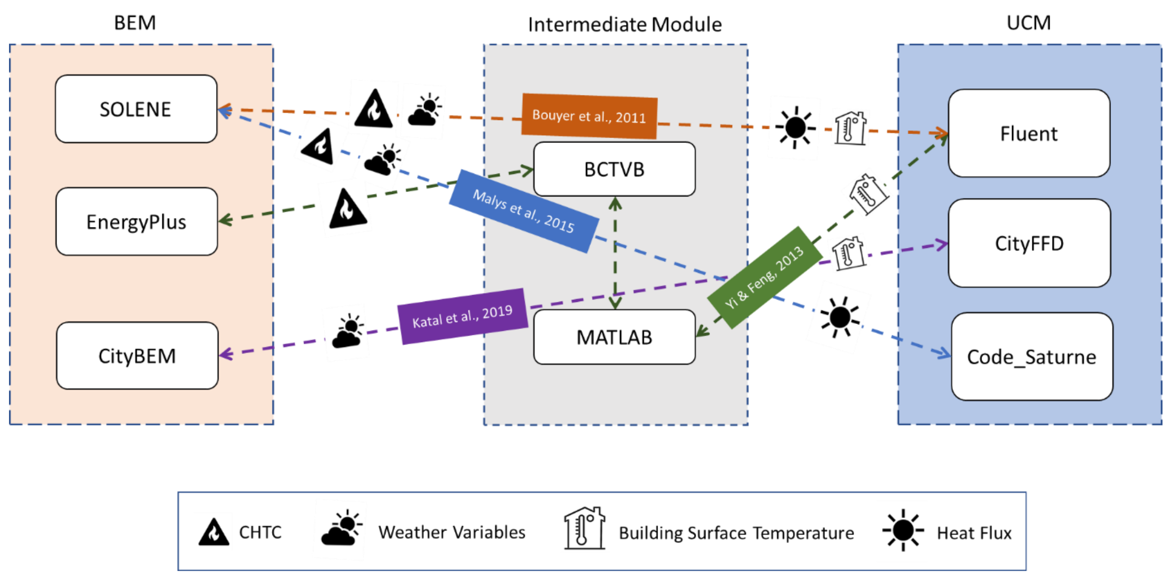


Figure 2-6. Two Directional Coupling Methods.

The resilience of buildings to extreme weather was investigated using a two-directional coupling method between CityFFD and CityBEM software Average atmospheric variables and convective heat transfer coefficients were exchanged between the models, enabling simulations of heating and cooling load and indoor air temperature, etc. Non-geometric input data was processed by CityBEM, while building surface temperature calculated by CityBEM was passed onto the CityFFD as the boundary conditions. This iterative data exchange was referred to as a 'ping-pong' data exchange method. To initiate this mechanism, the value of the building surface temperature was required as the first-time step. To fulfill that, a loop of 0.1% error condition was run with an assumed building surface temperature. It was compared with the final building surface temperature obtained by the CityBEM simulation. This converged value was used as the initial parameter for the CityFFD simulation.

Bouyer et al. improved building energy performance by coupling BES and CFD using a two-directional dynamic method [32]. However, fully dynamic was computationally expensive, so a low-resolution quasi-dynamic approach was used for turbulence, momentum, continuity, etc. Fluent and SOLENE exchanged variables such as air temperature, CHTC, and mass rate of moisture. The tool proved effective in understanding the influence of urbanization on building energy demands at a small scale. Malys et al. performed a sensitivity analysis using the coupled software SOLENE-Microclimate [33], following a similar approach [32]. CFD coupling had a limited impact on the winter season but significantly affected buildings with inadequate thermal insulation during the summer.

An overview of BES-CFD coupling methods was conducted to identify untested approaches and validate them [34]. The literature revealed limited studies on BES-CFD coupling, particularly for outdoor conditions. The focus was on the external wall as a reference for variable sharing. Yi and Malkawi's [35]outdoor integration method, untested in previous research, was explored using a "ping pong" strategy. A dynamic simulation module called BCVTB was incorporated, where BES provided building surface temperature to CFD and CFD provided CHTC to BES. MATLAB served as an intermediate for BCVTB and Fluent coupling. To ensure result accuracy, a CHTC deciding criterion was established based on literature-derived empirical models. If CFD-calculated CHTC failed the criterion, empirical model values were used. However, further research on outdoor conditions is urged. MoWiTT, TARP, and DOE-2 were employed to evaluate the criterion, but CHTC values from CFD were successfully used throughout the entire simulation.

2.5 Machine learning-based building energy models

In recent years, ML techniques have gained increasing attention for building energy prediction due to their ability to capture complex relationships between inputs and outputs. However, the development of robust ML models is often hindered by limited access to high-quality measured data, especially in scenarios where data privacy is a concern, or the building is still in the design or construction phase. To overcome these challenges, researchers applied physics-based simulation engines to generate synthetic building energy datasets. These synthetic datasets enable the creation of large, diverse, and labeled training data under controlled variations of building geometry, systems, usage profiles, and weather conditions.

ML techniques offer the flexibility to incorporate vast amounts of data gathered from various sources for predicting new samples and leading to informed decisions. Data sources could be smart grids, sensor networks, and on-site measurements, among others. An occupant-behavior-sensitive modeling approach was applied to four ML approaches for predicting building energy consumption [36]. Synthetic data generated by EnergyPlus serves as training and testing data resource, including 3-month hourly data. Li et al. [37] utilized two different historical data sets in hourly intervals, which are four months and three months in total to predict a building's electricity consumption using optimized artificial neural networks (ANN). Platon et al. [38] collect 14 months for 23 variables in the hourly time step, using the ANN model to predict hourly electricity consumption. XGBoost consistently demonstrates strong predictive performance across various building energy applications. For example, Wang et al. [39] employed multiple ML models to predict a detailed cooling load profile one hour ahead. The study concluded with a recommendation to utilize the XGBoost algorithm as they produced the best performance on both the training and testing set of the data. Similarly, among 17 ML models, XGBoost achieved the best accuracy in predicting solar radiation on urban building surfaces [40]. Wu et al. [41] also observed that XGBoost excelled in developing baseline energy models across various building types and time intervals among 8 different ML models. Ma et al. [42] demonstrated XGBoost's outstanding performance in predicting outcomes for both envelope and heating, ventilation, and air conditioning (HVAC) system retrofit strategies.

2.6 Limitation

Coupling strategies used for exchanging information between different simulation tools have several limitations. One common strategy, BEM+CFD, which simulates the convective heat transfer between buildings and surrounding urban areas, faces challenges due to the high computational cost and the need for powerful computing devices for CFD simulations. Furthermore, CFD simulations are highly sensitive to mesh properties, requiring high-quality meshing for accurate results, but this increases the simulation time. Implementing CFD tools also demands extensive knowledge and experience in simulation, as the lack thereof can result in less accurate results. Another limitation lies in the use of BEMs, which have constraints in incorporating exterior variable boundary parameters. Additionally, the time scales between BEM and UCM (Urban Canopy Model) often do not align, posing challenges for their coupling. There is a need to further develop BEM software to include all external heat fluxes, especially longwave radiation. Currently, using a specific coupling method for different scenarios can lead to complications. Microclimate studies using CFD analysis may produce local microclimatic conditions that deviate from reality. Uncertainties in case studies should be carefully considered when calibrating CFD-BES (Building Energy Simulation) models for varying scenarios. Most coupling strategies also overlook important factors such as the evapotranspiration of vegetation and evaporative cooling effects of water bodies, whereas ENVI-met takes these effects into account. Therefore, there is room for improvement and addressing these limitations in future coupling strategies.

2.7 Summary

This work provides an updated review of coupling strategies between urban microclimate and building energy models. It discusses available modeling tools, coupling methods, and recent research progress. Coupling urban microclimate and building energy simulations requires careful consideration for accuracy. Guidelines recommend one-directional or dynamic coupling methods, with static being easier but dynamic requiring more resources. Quasi-dynamic and fully-dynamic methods yield similar results unless airflow parameters change significantly. Coupling is less crucial for winter conditions but important for summers, especially in well-insulated buildings. The intermediate coupling method offers a good compromise. Common variables for CFD-BES coupling are building surface temperatures and meteorological data. Calibration involves using accurate data and adjusting the models based on experimental results.

Coupling urban microclimate and building energy simulations is crucial for accurate results. One-directional coupling is commonly used due to computational challenges, and there is a lack of research on coupling platforms. However, more studies are expected in the future to explore different coupling schemes. Further research is needed to investigate coupling possibilities and promote adoption among researchers and industry professionals.

Chapter 3. Assessment of urban microclimate and its impact on individual building energy performance

This chapter examines the urban microclimate on building energy performance in Qatar, a hot-arid climate, using both measurement data and computational modeling. This study collects measurement data across Qatar and conducts computational fluid dynamics (CFD) simulations; the results from both methods serve as inputs in building energy simulation (BES). The results demonstrate that space cooling demand is more sensitive to ambient temperature than other climatic parameters, building thermal properties, etc. The UHI intensity is high during hot and transition seasons and reaches a maximum of 13 ℃. BES results show a 10% increase in cooling energy demand for an office building due to the UHI effect on a hot day. The results of this study enable more informed decision-making during the building design process[2].

3.1 Introduction

Today, around 55% of the world's population lives in cities, and this rate is projected to reach as high as 70% by 2050 [1]. The State of Qatar stands out with the world's highest urbanization rate (99.1%), while its rural population represents only 0.9% [43]. Among different energy-consuming sectors, buildings account for over 40% of the annual energy consumption worldwide and around 55% of the world's electricity consumption [44], [45]. This rate is even higher in Qatar [46]. Thus, accurate estimation of building energy consumption is critical to provide city planners and policymakers with exquisite information on energy use to establish energy-efficient cities and mitigate greenhouse gas emissions and climate change. To this end, building energy simulation (BES) tools are widely used to estimate building energy consumption and investigate the influence

[2] This chapter has been published as a peer-reviewed journal paper: Dongxue Zhan, Nurettin Sezer, Danlin Hou, Liangzhu Wang, and Ibrahim Galal Hassan (2023). "Integrating Urban Heat Island Impact into Building Energy Assessment in a Hot-Arid City" *Buildings* 13, no. 7: 1818. https://doi.org/10.3390/buildings13071818

of input variables on the energy performance of buildings. However, typical meteorological Yyear (TMY) data have been widely used in BES studies to represent the ambient climate of the building area without considering the local urban heat island (UHI) effect, omitting complex interactions between buildings and the environment.

The UHI refers to the characteristic warmness of an urban area, which is often approximated by comparing the temperature of a city with its surrounding rural areas. Increasing urbanization aggravates the urban climate environment characterized by the UHI effect. The UHI intensity, a crucial indicator of the increased heat in urbanized areas, is defined as the air temperature difference between urban and rural areas [47]. Previous publications have investigated the UHI for various climates; for instance, the daily profile of urban air temperature was studied in comparison to that of rural areas for an observation period of one year in Switzerland [48]. The average diurnal UHI intensity varied from 0 ℃ to 2 ℃ and peaked at 10:00 pm on a sample clear-sky day on 26 June 2002. The UHI intensity reaches higher temperatures in cities with hot and arid climates. For instance, in the case of Doha, long-term measurement data show that UHI intensity reaches as high as 5 ℃ [49]. A lack of knowledge of the local UHI effect will decrease the accuracy of building energy simulation results. Shi et al. [50] presented the huge influence of the local UHI effect on the sensible and latent cooling energy demand of residential buildings during the summer in Hong Kong. The results show sensible cooling demand considering the urban microclimate is approximately twice that of the rural weather, and the latent cooling demand could be up to 96% higher. Heat in cities increases cooling demand, with each 1 °C increase causing an increase of between 8.0 and 15% in building use, placing a considerable burden on decarbonization efforts in cities [51]. In a bibliometric review of urban heat mitigation and adaptation, He et al. [52] summarized the impact assessment and cause identification of UHI. It has been suggested that a holistic and comprehensive understanding of the scope of urban heat and its associated impacts is needed.

A promising solution to investigate the impact of local UHI on building energy use is to couple urban microclimate simulation tools with BES tools since there is no distinct tool that can directly assess the urban microclimate impact on building energy use [7]. Urban microclimate simulation tools predict local ambient conditions regarding different urban configurations. However, the features and the thermal processes of buildings are usually simplified or neglected in these simulation tools. BESs can provide detailed descriptions of the building and its systems using a dynamic model in the building energy performance analysis involving many input parameters. There is a gap in explicitly quantifying the impact of local UHI on building energy consumption compared with other inputs in BES, especially in hot and arid climate zones. The specific objectives of this study are listed as follows:

- To provide insights into the significance of ambient temperature in BES compared to other building passive design parameters by global sensitivity analysis.
- To assess the tempo-spatial UHI effect with the help of year-round observation of six weather stations across the country in hot-arid coastal areas.
- To employ a one-way coupling strategy between urban microclimate simulation and building energy simulation using CityFFD and EnergyPlus, respectively.
- To demonstrate the impact of the UHI effect on building energy performance based on co-simulation results.

3.2 Description of Study Area

Over the last two decades, Qatar has experienced rapid population and urban growth [53], [54]. To accommodate part of Qatar's growing population, a new city development project, namely Lusail City, was developed as a flagship project to build a sustainable city on the northeast coast of Qatar. The project aimed to apply sustainability principles such as reducing greenhouse gas emissions, conserving water, and reversing desertification. Lusail City covers a 38 km^2 area and provides homes to more than 200,000 people. Marina district is a part of Lusail City. Its master plan comprises 36 high-rise commercial buildings, 37 mixed-use high-rise buildings, and 30 residential buildings to be constructed in 10 years. Until now, 20% of the buildings have been constructed, 27% are currently under construction, and 53% are still to be constructed. Pavement widths in the Marina district are determined by traffic volume, surface movement requirements, and under-grade infrastructure tunnel requirements. In order to enhance the sustainability of the district plan, greenery will be integrated into the rooftops, parking areas, and pedestrian areas of the district. To achieve the sustainability goals of the project, it is essential to evaluate the energy performance of buildings within the local urban context. BESs are commonly used to predict the space cooling and heating demands of buildings based on energy conservation for a control volume. Space cooling accounts for approximately 80% of the generated electricity in Qatar [55]. District cooling has the potential to save approximately 40% of the electricity due to the concentration effect of cooling load at one location [56], [57]. District cooling load calculations play a vital role in the design and operation phases of a district cooling plant. However, the cooling load of individual buildings in a district is usually overestimated in the BES modeling [58], which results in designing oversized district cooling systems, high initial investment, low operational efficiency, and waste of energy and water. Thus, accurate prediction of building cooling loads is necessary, yet remains a challenge in the optimal design and operation of district cooling systems.

The local environmental setting in Qatar underwent considerable changes as a result of the dynamic evolution of numerous urbanization aspects, including a general change in land use, the number of buildings, the total amount of living space, and the number of cars. Typically, these shifts take the form of potentially significant micro-level variances in the urban climate across the nation. Because of these factors, Qatar is an intriguing case study to examine the sensitivity of climate conditions to a range of urban-scale factors.

The Marina district is in the southern part of Lusail City, comprising high-rise towers (number of floors ranging from 15 to 40) for office, residential, mixed-use, hotel, and retail use connected to a continuous boardwalk (Figure 3-1). The district totals 3.1 million m^2 of built-up area, and the population of this district is 40,760, of which 27,000 are residents. It is bounded in the south by the Lagoons Canal, to the west by Road B, to the east by the Arabian Gulf shoreline, and to the north by Qatar Entertainment City. The urban geometrical data were extracted from the information provided by the utility representatives and through the literature search. The Marina district is projected to be the future downtown of Lusail City in Qatar, which is under colossal construction. Our case study focuses on a typical high-rise office building in the Marina district of Lusail City, Qatar. The selection was made based on the fact that all office buildings in the Marina district adhere to the same construction requirements outlined in local and national standards. In addition, the availability of relevant information regarding the geometry and features of this building contributed to the decision to choose this building.

3.3 Methodology

This section outlines the methods employed in the present work. Firstly, high temporal granularity weather data were collected from a locally installed weather station and five rural airport weather stations across the country. Additionally, the microclimate was assessed using CityFFD, a microclimate tool based on CFD simulations. The coupling of building energy simulation and microclimate modeling enabled the analysis of UHI effects on building energy performance. In conducting a global sensitivity analysis, key climatic variables and building design features that have a substantial impact on building energy modeling were identified.

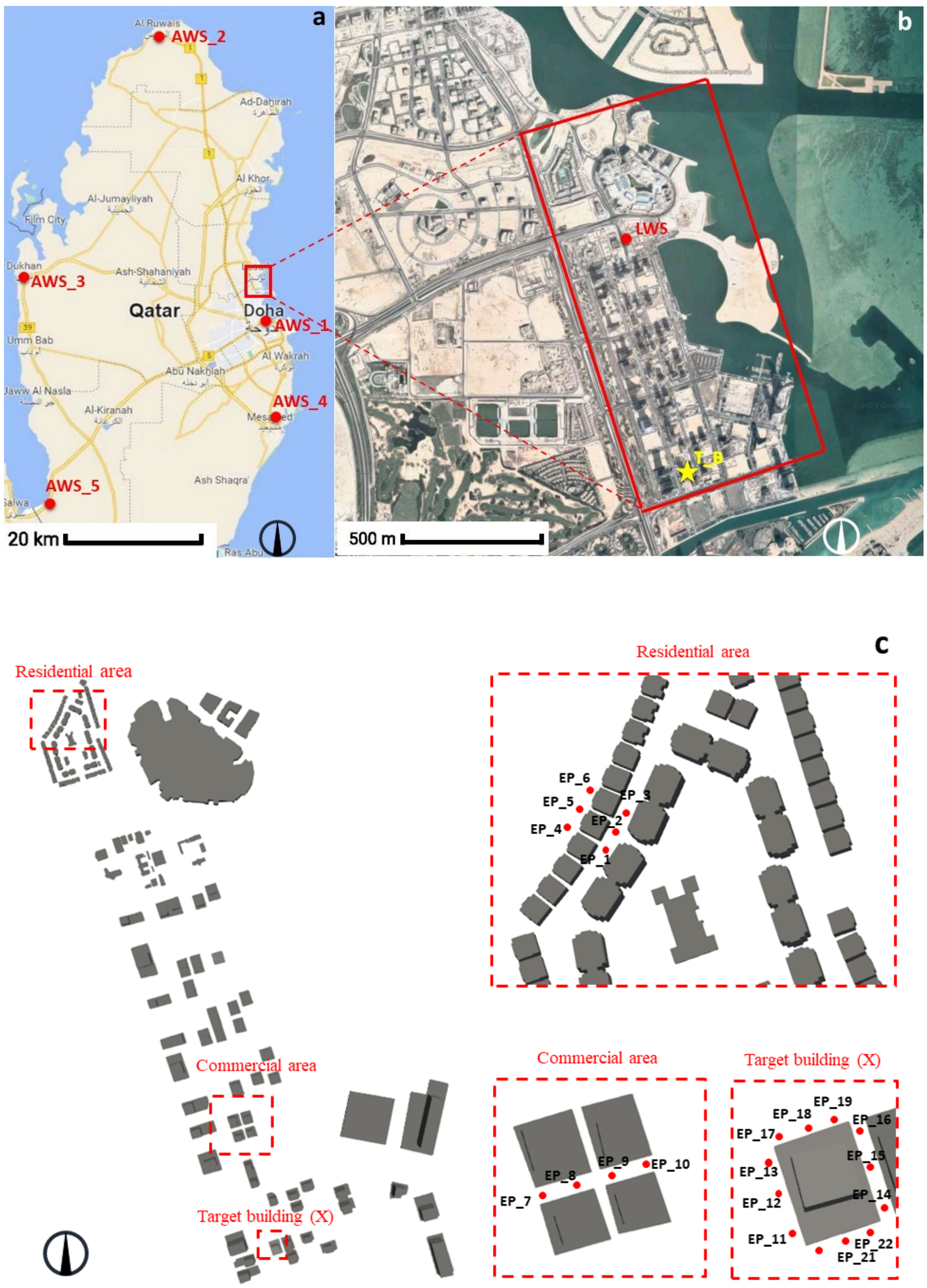


Figure 3-1. (a) Locations of airport weather stations in Qatar, (b) urban layout of the Marina district, and (c) locations of evaluating points. (AWS: airport weather station; LWS: local weather station; T_B: target building; EP: evaluating point).

3.3.1 Weather data collection

High temporal resolution meteorological information is acquired from a local weather station set up in the Marina district of Lusail City at coordinates 25.399952 E and 51.519568 N. The sensors of the weather station report hourly data on temperature, solar radiation, wind direction, wind speed, relative humidity, and precipitation. Wind speed and temperature sensors were positioned 5 m and 4 m above the ground, respectively, and the measuring pole was 8 m away from the building.

In addition to the local weather station installed in the Marina district of Lusail City, a total of five airport weather stations were installed in different locations across Qatar: Doha International Airport, AI Ruwais, Dukhan, Umm Said, and Abu Samra. These have been designated as AWS-1 to AWS-5 in this study. The location of each station is shown in Figure 3-1. The airport meteorological information was collected from the OGIMET website based on the "climate" package in R 3.6.3 [59], which specializes in the automation of meteorological data downloading. Hourly meteorological data are available for these stations during summer, while every three-hour data are available during winter. UTC +3 time zone is used in simulations and the time difference between the airport weather station and local measurement is accounted for. The five airport meteorological stations were installed in different locations across Qatar. All aforementioned weather data were collected from August 2020 to August 2021.

3.3.2 Urban microclimate modeling

Modeling urban scale microclimate and capturing neighborhood building impact on the aerodynamics often lead to huge computing loads. CityFFD is based on the semi-Lagrangian approach, a fast and stable numerical model suitable for modeling large-scale airflow problems [60]. The turbulence closure is achieved by the large eddy simulation (LES) [61]. CityFFD solves the following non-dimensional Navier-Stokes conservation equations [61]:

$$\nabla \cdot U = 0 \quad \text{Equation 3-1}$$

$$\frac{\partial U}{\partial t} + (U \cdot \nabla)U = -\nabla p + \left(\frac{1}{Re} + v_t\right)\nabla^2 U - \frac{Gr}{Re^2}T \quad \text{Equation 3-2}$$

$$\frac{\partial T}{\partial t} + (U \cdot \nabla)T = \left(\frac{1}{R_e \cdot Pr} + \alpha_t\right)\nabla^2 T + q \quad \text{Equation 3-3}$$

where t is dimensionless time, U is dimensionless velocity, p is dimensionless pressure, T is dimensionless temperature, and q is the dimensionless heat source.

CityFFD was already validated in our previous studies by the numerical and experimental data in the literature [62], [63]. In this case, the computational domain size was 4000 × 4900 × 1110 m^3. A grid convergence study was performed, the select meshing setting shows a 1 m grid size close to the building. The total grid number is 65.7 million. A 24 h period was simulated with 1 h timesteps and 4000 iterations per timestep. The turbulence closure was achieved by the large eddy simulation (LES) [61].

3.3.3 Building energy modeling

EnergyPlus calculates the local outdoor wind speed and air temperature separately for each zone, as well as the temperature of surfaces exposed to the outdoor environment according to the US Standard Atmosphere [64] and Handbook of Fundamentals (ASHRAE 2005) [65]. However, the variation in barometric pressure is ignored in most situations [66].

There is a strong correlation between altitude and air temperature. Air temperature decreases almost linearly with altitude at a rate of ~1 °C per 150 m across the troposphere. Wind speed increases with altitude whereas barometric pressure gradually decreases with altitude. Therefore, tall buildings could experience significant differences in local atmospheric conditions between the ground floor and the top floor [66], [67].

Variation in outdoor air temperature is calculated using reference values for a given altitude independent of climate and seasonal differences [66], [67], based on the US Standard Atmosphere (1976) [64]. The following formulas describe the relationship between air temperature and altitude in any given layer of the atmosphere.

$$T_z = T_b + L(H_z - H_b) \qquad \text{Equation 3-4}$$

$$H_Z = \frac{Ez}{(E + Z)} \qquad \text{Equation 3-5}$$

$$T_b = T_{z,met} - L\left(\frac{E_{Z_{met}}}{E + Z_{met}} - H_b\right) \qquad \text{Equation 3-6}$$

where T_z and T_b are the air temperature at altitude z and the base of the layer, i.e., ground level for the troposphere, respectively; L is the air temperature gradient ($L = -0.0065\ \mathrm{K/m}$ in the troposphere); H_b is offset equal to zero for the troposphere; H_z is geopotential altitude; E is the radius of the Earth, which equals 6356 km; Z is altitude; $T_{z,met}$ is weather file air temperature (measured at the meteorological station); and Z_{met} is the height above the ground of the air temperature sensor at the meteorological station. The default value for Z_{met} for air temperature measurement is 1.5 m above the ground level.

The local wind speed calculations were performed as described in Chapter 16 of the Handbook of Fundamentals (ASHRAE 2005) [65]. The wind speed measured at a meteorological station is extrapolated to another altitude with the following equation:

$$V_z = V_{met}\left(\frac{\delta_{met}}{z_{met}}\right)^{\alpha_{met}}\left(\frac{z}{\delta}\right)^{\alpha} \qquad \text{Equation 3-7}$$

where z is altitude (the height above ground); V_z is the wind speed at altitude z; α is the wind speed profile exponent at the site; δ is the wind speed profile boundary layer thickness at the site; z_{met}

is the height above ground of the wind speed sensor at the meteorological station; V_{met} is wind speed measured at the meteorological station; α_{met} is the wind speed profile exponent at the meteorological station; and δ_{met} is wind speed profile boundary layer thickness at the meteorological station. The wind speed profile coefficients, α, δ, α_{met}, and δ_{met}, are variables that depend on the roughness characteristics of the surrounding terrain.

To investigate the impact of the UHI effect on building energy consumption, a detailed commercial BES model was developed using OpenStudio, SketchUp, and EnergyPlus, as shown in Figure 3-2. The target building was Sendian Tower, which is a typical high-rise commercial building in the Marina District of Lusail City, Qatar. It comprises 2 basement levels, a ground-floor lobby lounge, 27 additional floors, and a penthouse. For the difficulty of gathering some parameters, realistic assumptions were made during the modeling of the target building. A total of 31 thermal zones are defined in the model, with the assumption that each floor corresponds to a single thermal zone. While this simplifies the simulation, the single-zone-per-floor approach limits the ability to capture intra-floor thermal variations, diverse occupancy patterns, and localized HVAC control, potentially affecting the accuracy of energy performance predictions. The heating, ventilation, and air conditioning (HVAC) system for each zone is a four-pipe fan coil system for space cooling. People density is 0.057 person/m^2. Lighting power density is 10.65 W/m^2, and electricity equipment power density is 7.64 W/m^2. The timestep is 10 min, which serves as the driving timestep for heat transfer and load calculations in the zone heat balance model. Table 3-1 summarizes the characteristics of Sendian Tower.

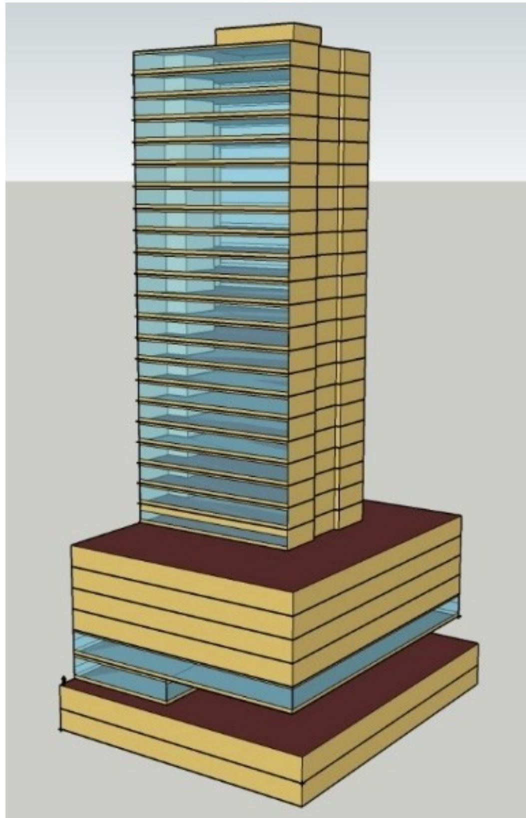


Figure 3-2. Southeast face of the office building model in EnergyPlus

Table 3-1. Target building characteristics

Archetype/Feature	**Description**
Location	The Marina district of Lusail City, Qatar.
Building type	Commercial building
Year of construction	2017
Number of floors	Two basements, 1 ground floor, 27 typical floors, and a penthouse.
Total building area	41,619.09 m^2
Number of thermal zones	31
HVAC system	The four-pipe fan coil system
Thermostat setting	24 °C from 6:00 to 22:00 and 26.7 °C for the rest of the day.
People density	0.057 person/m^2
Lighting power density	10.65 W/m^2
Electricity power density	7.64 W/m^2

In the present version of EnergyPlus, air temperature, wind direction, and wind speed are obtained from the TMY file, a meteorological file, and assumed identical for all the surfaces in the scene. For the link to the urban microclimate model, the EnergyPlus model is adapted, allowing for the allocation of individual outdoor air temperature and wind speed to each surface element of the building envelope.

3.3.4 Coupling strategy

A one-way coupling strategy was used to perform a co-simulation of CFD and BES. A simplified simulation framework is illustrated in Figure 3-3. The CFD model calculates local microclimate based on airport meteorological data and urban geometry. The airport meteorological data includes hourly air temperature, wind speed, wind direction, and humidity, from August 2020 to August 2021. The site-specific air temperature around the target building was extracted from the CFD results and then integrated into the EPW file as EPW_{local} for EnergyPlus meteorological boundary condition, taking into account the urban microclimate impact. With the help of SketchUp and OpenStudio models, the building cooling load and surface temperature were calculated in EnergyPlus.

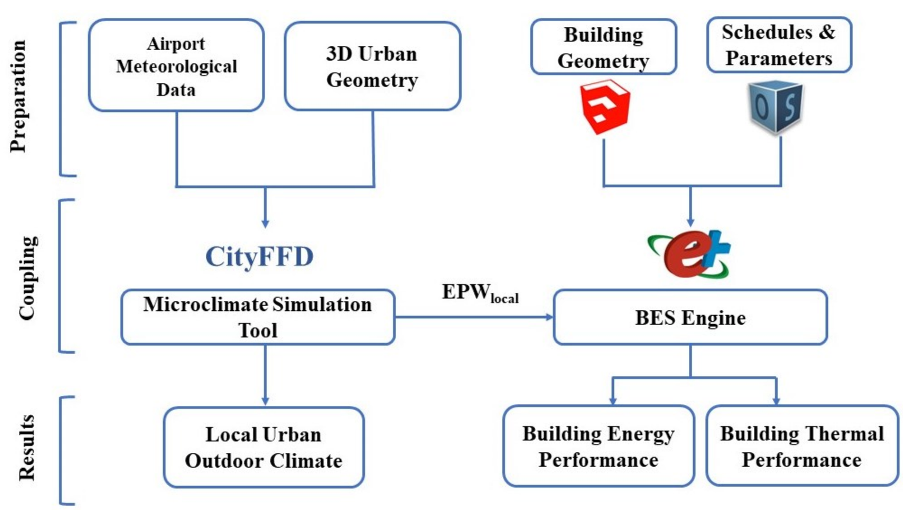


Figure 3-3. One- way framework of the co-simulation of CityFFD and EnergyPlus.

3.3.5 Sensitivity analysis

A global sensitivity analysis was conducted to consider the influences of uncertain cases over the whole input space in this study. A typical SA study consists of six steps as follows: (1) determining input variations, (2) creating building energy models, (3) running energy models, (4) collecting simulation results, (5) running sensitivity analysis, and finally, (6) presenting sensitivity analysis results [68]. Besides the ambient air temperature, building energy consumption was influenced by many input parameters, including weather data, thermal properties of the envelope, and internal gains. The building envelope parameters and their variations are defined by ASHRAE or Qatar local standards (Lusail City GSAS 2 Star Rating Guidelines) [69]. In total, 13 input parameters were evaluated in this study; the range and data source are summarized in Table 3-2. The Latin Hypercube Sampling (LHS) method [70], using the R "lhs" package, was applied due to providing good convergence of parameters space with relatively few simples. A total of 605 different combinations of the parameters were utilized as inputs in this work. A parametric simulation was conducted using the EnergyPlus model to generate hourly and annual cooling loads using the RStudio script. The R "eplusr" package [71] was used to perform the parametric simulation and automatically collect input and output datasets.

Table 3-2. Input parameters and ranges of values.

Number	Parameters	Unit	Range of Values	Source
1	Air temperature	°C	8.9–46	Doha TMY weather data
2	Wind speed	m/s	0–25.7	Doha TMY weather data
3	Cooling set point	°C	21–26	[69]
4	Lighting power density	W/m^2	0–9	[69]
5	Relative humidity	%	5–100	Doha TMY weather data
6	Wall insulation U-value	$W/(m^2K)$	0–0.3	[69], [72]

7	Infiltration	ACH	0.1–0.2	[72]
8	Ventilation	m^3/s/person	0.00047–0.00247	[72]
9	Occupancy density	m^2/person	19–24	[72]
10	Solar reflectance of interior diffusing blinds roll	/	0.4–0.8	[73]
11	Roof insulation U-value	W/(m^2K)	0–0.25	[74]
12	Window solar heat gain coefficient	/	0–0.22	[72]
13	Window U-value	W/(m^2K)	0–1.8	[72]

This study applied a sensitivity analysis (SA) method and sensitivity value index (SVI) method [68], [75] to compensate for the difference in sensitivity analysis methods and target output. The SVI method is integrated with the standardized regression coefficient (SRC), random forest variable importance, and T-value method. The importance ranking among input parameters for the target high-rise office building was identified using the SA approach. The SVI calculation is performed based on the following formula [44]:

$$SVI\ (\%) = \sum_{l=1}^{m} \frac{\frac{\sum_{j=1}^{k}\left(\frac{V_{i,j}}{\sum_{i=1}^{n}\left|V_{i,j}\right|}\right)}{k}}{m \cdot k} \times 100 \qquad \text{Equation 3-8}$$

where v is the value of a sensitivity analysis method, i is a parameter, n is the total number of the parameters, j is a sensitivity method, k is the total number of sensitivity methods ($k = 3$), l is the target output, and m is the total number of target outputs (m = 1; building cooling demand).

3.4 Results and Discussion

We present our findings in the conclusion section. The global sensitivity analysis conducted on a typical office building in the study area revealed significant variables that influence building energy modeling. Further, we demonstrated variations in ambient temperature and UHI intensity around the target building. Through analysis of UHI effects on building cooling energy loads, specifically on a typical hot day, we gained insight into the importance of considering UHI effects in building energy modeling. These findings collectively contribute to a better understanding of the factors influencing building energy performance. They emphasize the need to account for UHI effects in sustainable building design.

3.4.1 Global sensitivity analysis

The global SA of a typical high-rise office building was conducted as described in Section 4.5. The SA results are listed in Table 3-3. The impact of 13 input parameters was ranked from 1 to 13 according to their SVI values. The most critical input parameter is indicated as 1 and the least as 13. The results show that two model weather input parameters—air temperature, and wind speed—were the most critical parameters with the highest SVI values, followed by the cooling set point and lighting power density. However, solar irradiation plays the most crucial role in the energy model simulation. It was observed that most of the surrounding buildings are far away from the studied high-rise office building, thus the shading caused by other buildings has a slight impact on the target building. Therefore, we did not consider the solar-related parameters in the co-simulation. The conduction transfer function (CTF) thermal model, used in this study, was the best choice for the analysis of building energy simulation, especially in a hot climate. Notwithstanding, it does not account for the combined transport of heat and moisture within building envelopes [77]. Thus, the CTF model probably underestimates the relative humidity (RH) impact on building energy simulation.

Table 3-3. Results of sensitivity analysis

Parameters	SRC	Random Forest	T-Value	SVI	Rank
Air temperature	0.53	204.71	68.95	24.09	1
Wind speed	0.16	58.74	23.36	7.53	2
Cooling set point	0.13	60.93	20.97	6.85	3
Lighting power density	0.10	42.88	16.39	5.15	4
Relative humidity	0.01	58.41	0.66	2.56	5
Wall insulation thickness	0.02	7.08	3.67	1.05	6
Infiltration	0.02	6.50	3.74	1.04	7
Ventilation	0.01	3.97	1.33	0.44	8
Occupant density	0.01	6.06	0.90	0.43	9
Solar reflectance of interior diffusing blinds rolls	0.01	4.217	0.81	0.34	10
Roof insulation thickness	0.01	3.52	0.85	0.32	11
Window solar heat gain coefficient	0.003	2.694	0.52	0.22	12
Window insulation	0.0001	2.899	0.02	0.12	13

3.4.2 Urban heat island effect

Weather in Qatar is tropical maritime with two main seasons: a cold season from December to February and a hot season from May to October, with March, April, and November being the transition months [78]. In the study area, the Marina district consists of 36 high-rise commercial buildings, 37 mixed-use high-rise buildings, and 30 residential buildings. Figure 3-1 illustrates the layout of the buildings in the study area. The UHI intensity was estimated by calculating the difference between the air temperature of the local and rural stations. Based on the observed data from the local weather station and each airport weather station across Qatar, the maximum station difference between the air temperature of the local weather station and each airport weather during each season was calculated and plotted in Table 3-4 during the period between August 2020 and

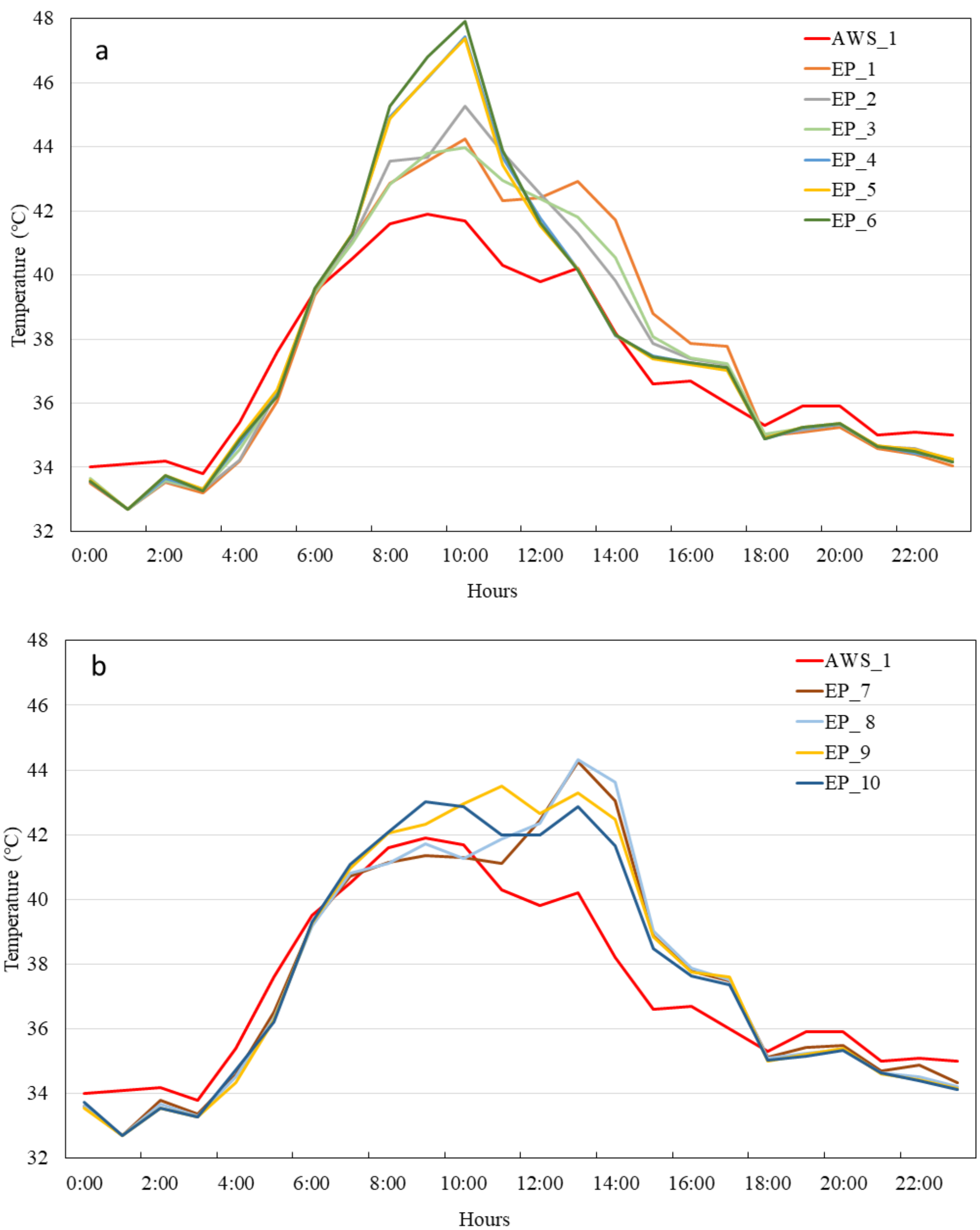
August 2021. It can be seen in Table 4 that the maximum air temperature difference in each season varies from 5.2 °C to 13.07 °C. The significant UHI effect occurs in hot and transition seasons, and a noticeable difference in wind speed was observed in the transition season. Different spatial distribution characteristics of rural air temperature were observed. The air temperature data collected from AWS_5, situated southwest of the local weather station, was higher during cold and transition seasons. The minimum air temperature difference was observed to be between the local weather station and AWS_1 due to a short distance. The results indicate that the selection of rural station data influences the UHI intensity.

Table 3-4. Maximum differences between the air temperature of the local weather station and each airport weather station during each season.

	AWS_1	**AWS_2**	**AWS_3**	**AWS_4**	**AWS_5**
Cold season	5.67	7.87	8.17	7.93	9.65
Hot season	5.20	11.58	10.30	11.99	9.50
Transition	6.39	11.41	11.81	7.71	13.07
All seasons	6.39	11.58	11.81	11.99	13.07

In order to evaluate the UHI intensity in Marina District, 22 evaluating points are selected to present the temperature variations. These locations include the high-density residential building area, high-density commercial building area, as well as near-wall of the target building. The locations of evaluating points are illustrated in Figure 3-1.

The hourly temperature variations of the evaluating points against the temperature data of AWS1 on a hot day are presented in Figure 3-4. As mentioned earlier, the UTC +3 time zone was used in the simulations. The peak air temperature occurred at different locations and times, such as 47.9 °C at 10:00, 44.3 °C at 13:00, and 46.4 °C at 13:00, in the residential area, commercial area, and target building, respectively. Here, the peak temperature is recorded as the highest value in the temperature data obtained from all evaluation points in each observation area. In addition, the average peak temperature of the evaluation points is calculated as 46.0 ℃, 43.8 ℃, and 45.6 ℃ in the residential area, commercial area, and target building, respectively. On the other hand, the peak temperature of AWS1 was measured as 41.9 ℃ at 9:00. All peak temperatures of simulated evaluating points are greater than the AWS1. These results indicate the spatiotemporal variation of the thermal intensity across the Marina district due to a number of factors such as building density, building envelope material and its thermal properties, and location of building stocks.



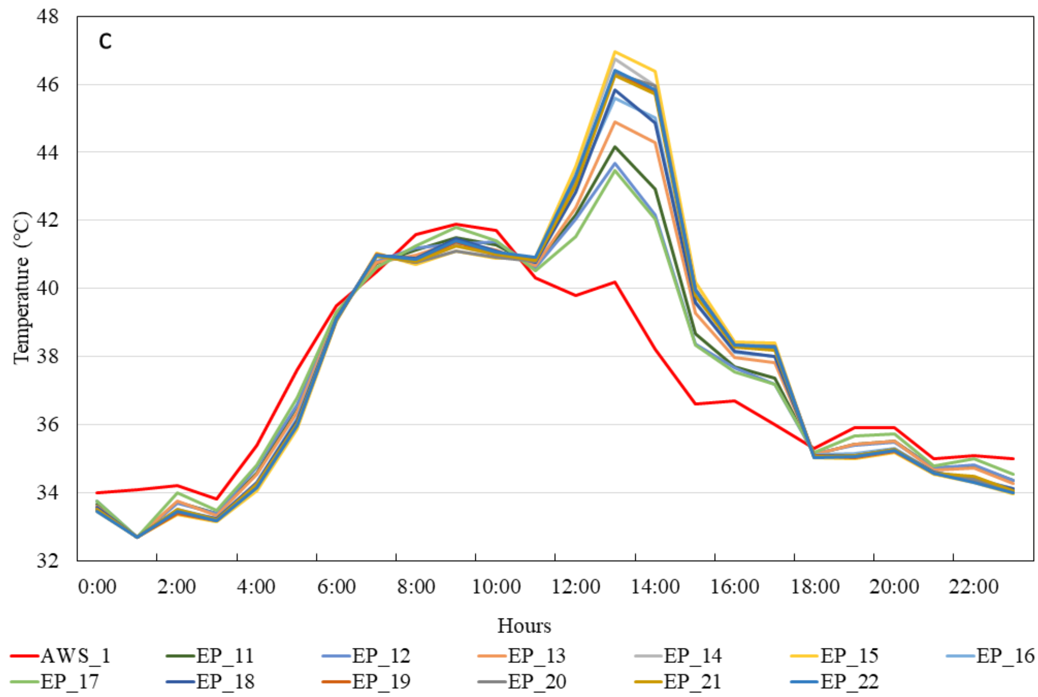
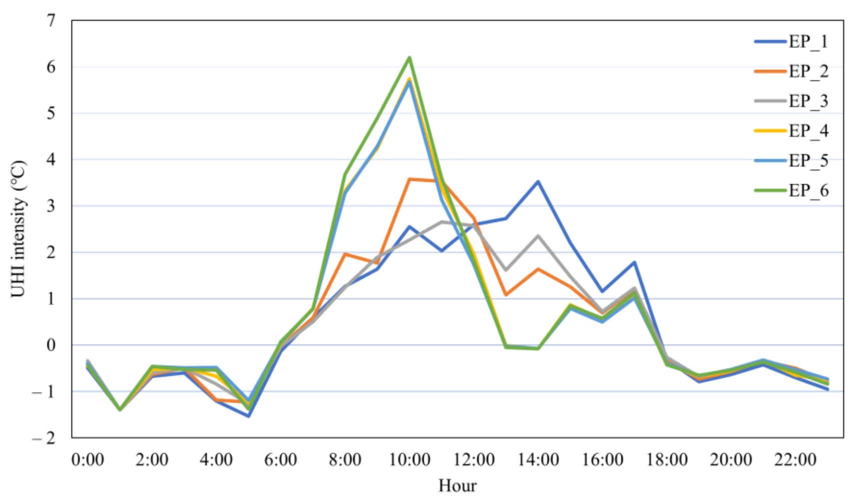


Figure 3-4. Temperature variation of evaluating points against the temperature data of airport weather station 1 on a hot day: (a) residential area, (b) commercial area, and (c) target building.

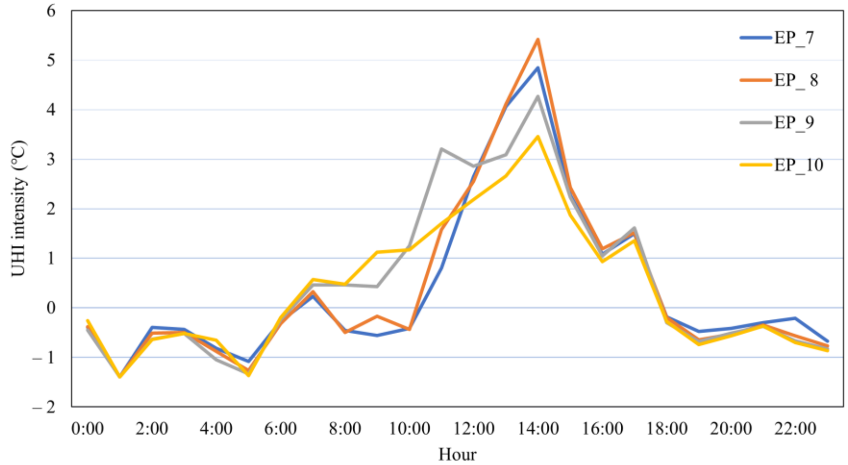
The spatiotemporal variation of the UHI intensity within the Marina district on a hot day is analyzed by plotting UHI intensity graphs based on the variation of the temperature difference between the simulated evaluation point and measured data from AWS1 by time (

Figure 3-5). The UHI effect is observed only in the daytime between 6:00 and 18:00. The UHI intensity reaches a maximum of 6.2 °C at 10:00, 5.4 °C at 14:00, and 8.2 °C at 14:00 in the residential area, commercial area, and target building, respectively. The average maximum UHI peaks are 4.6 °C, 4.5 °C, and 6.5 °C in the residential area, commercial area, and target building, respectively. Variation of UHI intensity among the simulation areas is due to the type of building stocks (building archetype) and location. The residential area, for example, is closer to the inlet airflow due to its location where the wind provides more cooling effect, resulting in lesser UHI intensity in this area. The peak temperature and UHI intensity of each area are summarized in Table 3-5.

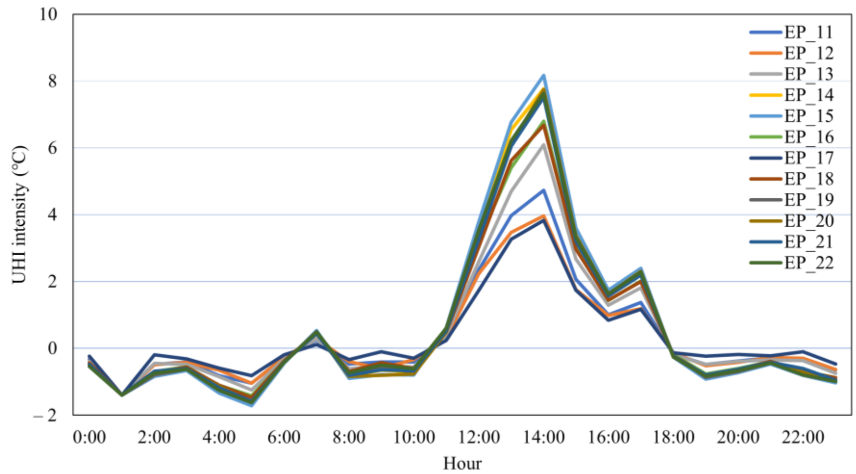
Mean absolute error (MAE) represents the average absolute difference between simulation results and corresponding measurement values over the dataset. It provides a linear score by assigning equal weights to all the individual differences in the average. On the other hand, root mean squared error (RMSE) involves squaring each difference between simulation and measurement data, averaging them, and taking the square root of the average. As a result, the RMSE emphasizes large errors by assigning relatively greater weight to them. The MAE and the RMSE were calculated together to diagnose the variation in the UHI intensity in the prediction values of the simulation on a hot day (Figure 3-6). The higher difference was obtained in the daytime between 8:00 and 15:00, with both errors peaking at 14:00, which indicates the UHI intensity reaches the maximum at 14:00 in the Marina district.



(a)



(**b**)



(**c**)

Figure 3-5. Spatiotemporal variation of the UHI intensity on a hot day in the Marina district: (a) residential area, (b) commercial area, and (c) target building.

Table 3-5. Summary of the peak temperature and UHI intensity data on selected areas and around the target building in the Marina district.

	Temperature (°C)			**UHI intensity (°C)**		
	Time	max	avr max	Time	max	avr max
Residential	10:00	47.9	46.0	10:00	6.2	4.6
Commercial	13:00	44.3	43.8	14:00	5.4	4.5
Target	13:00	46.4	45.6	14:00	8.2	6.5

Time: UTC +3, max: maximum; avr max: average maximum; Residential: residential area; Commercial: commercial Area, Target: target building.

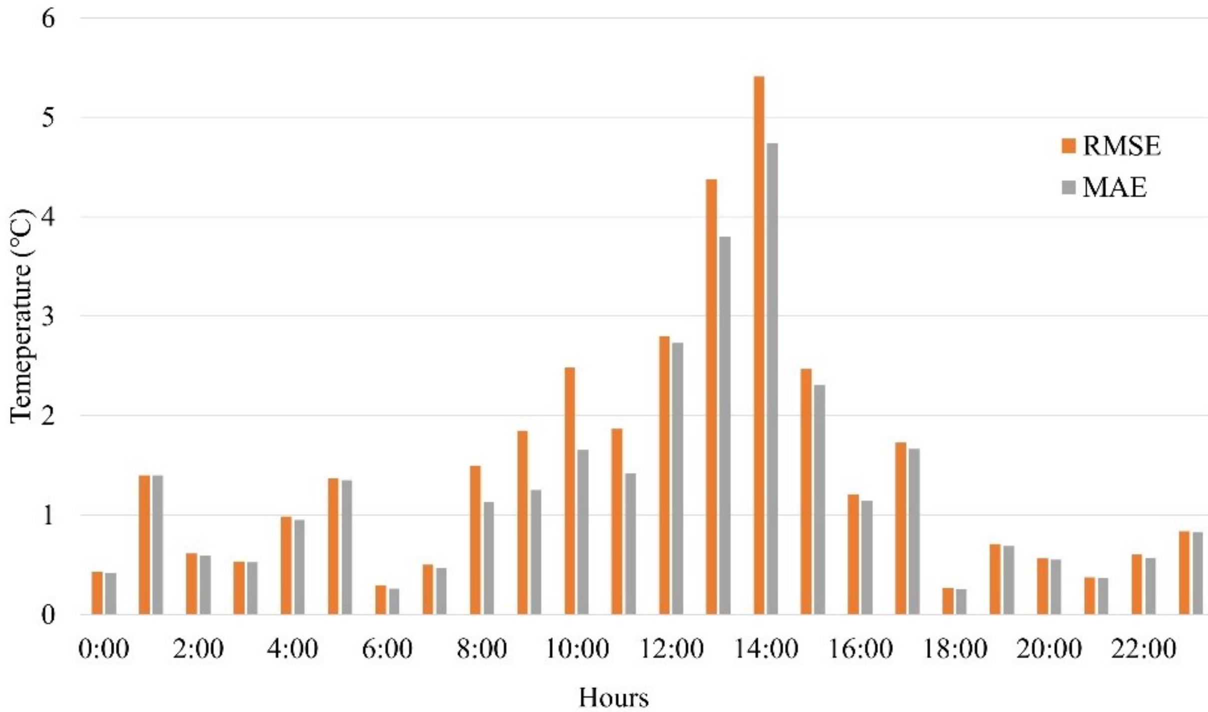


Figure 3-6. UHI intensity of evaluating points against airport weather station on a hot day.

3.4.3 UHI impact on cooling load

To evaluate the microclimate impact on building energy load, the date 9 August 2021, was selected as the evaluating day to represent a typical hot day. The temperature profile of the Marina district at 14:00 on a hot day was evaluated through CFD simulation (Figure 3-7). The wind direction from the north can be observed on temperature gradients adjacent to the buildings in the evaluation area. Building surface temperature varies among different building stocks, depending on many factors, including building characteristics, envelope material, and shading equipment use.

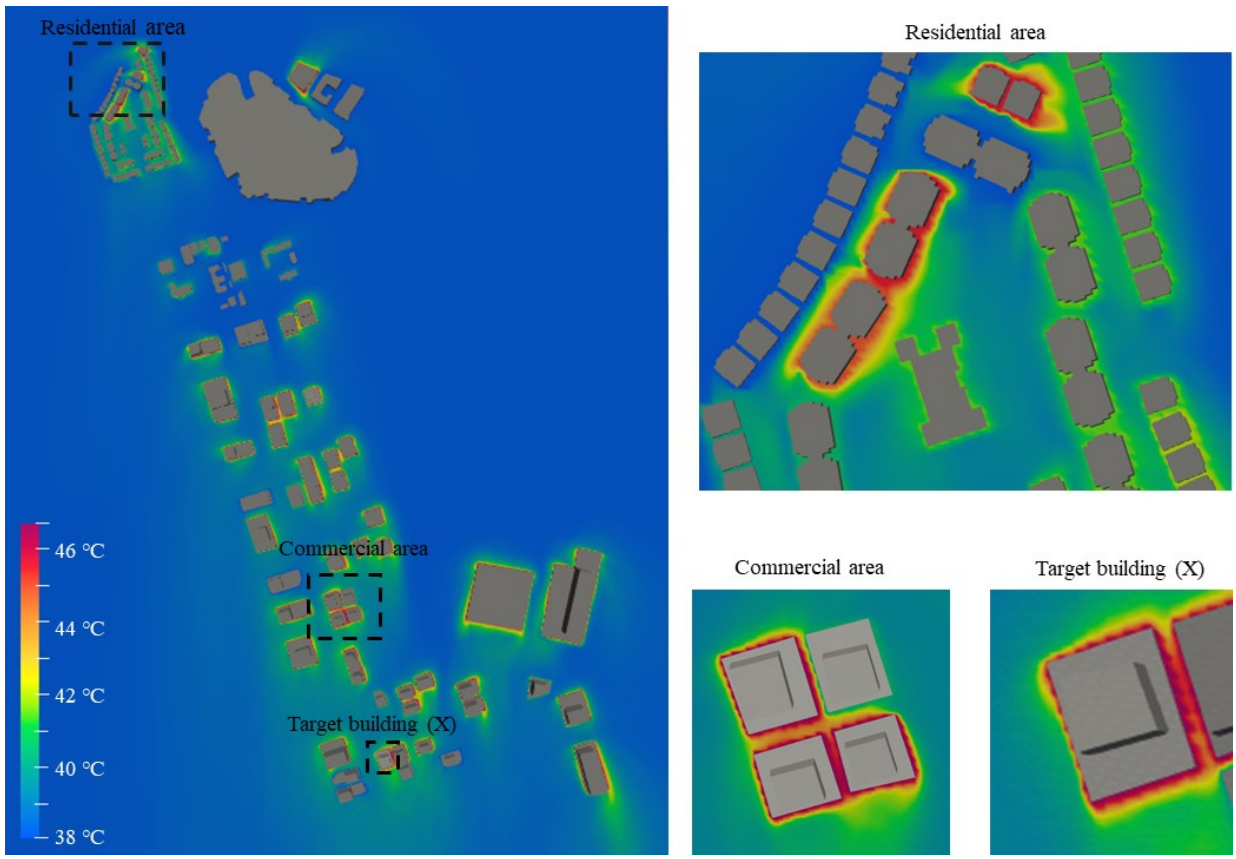


Figure 3-7. Thermal environment of the residential area, commercial area, and target building in the Marina district.

Urban microclimate impact on building cooling load on a hot day was investigated by inputting CFD results into EnergyPlus. The hourly cooling energy load intensity—the cooling energy used per unit floor area—was simulated in EnergyPlus. The average values of all evaluating points were considered as the local microclimate air temperature, which was integrated into the EPW file as EPW_{local} for following BES modeling taking into account the urban microclimate impact. The results are presented against the other simulation results with the data of each airport weather station in Figure 3-8. The minimum and maximum cooling loads were observed at 3:00 and 14:00, respectively. The daily cooling load obtained with the input of CityFFD was higher than that with data from each weather station. With the input of CityFFD results, the daily Energy Use Intensity (EUI) was computed as 0.53 kWh/m^2, whereas the average daily EUI based on the five weather station data was found as 0.48 kWh/m^2 (10% lower). The difference in cooling load was higher in the daytime than at nighttime due to the UHI effect in the daytime. With the input of CityFFD results, the daytime cooling load between 8:00 and 18:00 was computed as 11.9 MWh, whereas the average daytime cooling load based on the five weather station data was found as 10.4 MWh. The difference between results, based on CityFFD and each weather station data, indicates the microclimate impact on building energy consumption. The available literature related to the urban microclimate impact on building cooling energy consumption shows a wide variation among different studies, due to several factors such as different study areas, building characteristics, and the method of considering the urban microclimate impact.

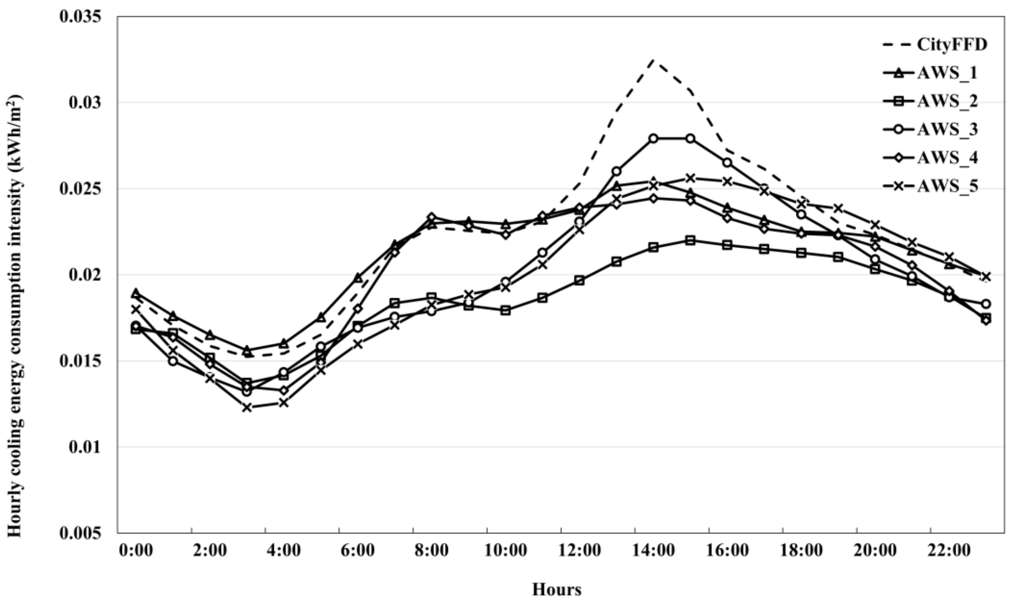


Figure 3-8. Building cooling load on a hot day.

3.5 Summary

This paper investigates the spatiotemporal characterization of an urban heat island (UHI) and its impact on building cooling load for a high-rise office building in the Marina district of Lusail City, Qatar. Global sensitivity analysis was conducted contributing to a better understanding of the urban ambient temperature impact on the building energy performance compared to other BES parameters. With the help of high temporal-resolution observed data collected from six weather stations across the country, the UHI effect in Qatar was analyzed representing the hot-arid coastal climate. CFD analysis was conducted to evaluate the UHI effect in the case study area. The CFD results obtained from CityFFD were inputted into EnergyPlus to simulate the building energy consumption. As such, one-way coupling enabled the evaluation of the impact of the UHI effect on building energy performance. The following conclusions can be drawn from this study.

- According to the impact ranking of input parameters in global sensitivity analysis, the most critical input parameter is the air temperature, followed by wind speed.
- The air temperature difference between the local weather station data and airport weather station data indicated the UHI effect of the urban area.
- The spatiotemporal variation of UHI intensity observed in residential and commercial areas in the Marina district stems from a number of factors, such as building density, thermal properties of building envelope material, shading equipment use, and the location of building stocks.
- The difference between MAE and RMSE results is minimal in the nighttime and maximum in the daytime, indicating the high UHI intensity during the daytime.
- The building cooling load obtained with the input of CityFFD results was higher than with weather station data. The difference clearly indicates the significance of considering the UHI impact in building energy simulation.

This study shows the significance of considering urban microclimate impact in BES studies. Coupling CFD and BES enables defining the meteorological boundary conditions accurately and obtaining realistic energy predictions of buildings within an urban context.

Chapter 4. Machine Learning-Based Hourly Building Energy Prediction Models

This chapter investigates the advantages of using the machine learning (ML) model as a surrogate model in building engineering, specifically for predicting hourly building energy consumption during the design phase. Synthetic data is commonly used for training and testing ML models when real-life measured data is unavailable due to privacy concerns or pre-construction scenarios. However, the challenge arises from the vast dataset of synthetic data generated by combining long-term hourly meteorological data with building characteristics. Using an example building, 82 million data points were generated as a result of simulating 8,760 hours when considering ten building performance parameters. To address this issue, a methodology utilizing weather clustering techniques is proposed in this work. This approach aims to reduce dataset size associated with day-by-day simulations by identifying representative weather patterns[3].

4.1 Introduction

Among various energy-consuming sectors, buildings use over 40% of annual energy worldwide and roughly 55% of electricity [2], [3]. As global warming increases and urbanization accelerates, building energy consumption is anticipated to increase. Energy efficiency improvement has become an essential aspect of reducing energy usage in buildings. Early design phases play an important role in determining building energy consumption [79]. For newly developed cities, the primary step in optimizing energy use is to evaluate energy consumption and identify trade-offs [80] by applying whole-building energy modeling. Building energy assessment on an hourly basis enables better support for effective energy management systems and day-ahead prediction [81].

[3] This chapter has been accepted as a peer reviewed journal paper: Dongxue Zhan, Shaoxiang Qin, Liangzhu (Leon) Wang, Ibrahim Galal Hassan (2025). "Weather clustering for machine learning-based hourly building energy prediction models at design phase." Energy and Buildings, https://www.sciencedirect.com/science/article/pii/S0378778825000386

Further, understanding hourly energy consumption is critical for estimating peak loads in a building, which are used to size equipment and design systems to handle peak demand to avoid oversizing/undersizing. It also improves the accuracy of monthly energy use, enhances system comparisons, provides more accurate load histories, and delivers higher-quality time-of-use energy data [82].

Building energy usage can be assessed through engineering calculations, simulation models, statistical models, and ML [83]. Engineering methods determine building energy consumption by using the laws of physics. A solid professional knowledge of complex mathematics or building dynamics is required to ensure results accuracy. Building energy efficiency simulation is used to simulate performance with predefined status. Building energy consumption can be calculated using a variety of simulation tools for efficient design and retrofits, both at the individual building level (DOE-2 [84], EnergyPlus [85], TRNSYS, e-Quest, etc.) and at the urban scale (IES VE, Eco-Tect, GreenBuilding Studio, CityBES [86], CityBEM [87], CitySim, and UMI). Most energy modeling software absorbs input parameters, involving weather data, building geometry, envelope thermal properties, operating schedule, and internal systems settings. Next, the software applies an engineering model to quickly estimate the energy consumption of a particular building. However, these tools are computationally intensive and require some expertise to operate them. Tasks include gathering building geometry and detailed information on buildings, constructing energy models, and performing calibration and validation. These statistical approaches use historical data and commonly employ regression techniques to model a building's energy consumption. Utilizing the provided data, ML uncovers relationships between input features and output targets. ML techniques offer the flexibility to incorporate vast amounts of data gathered from various sources for predicting new samples and leading to informed decisions.

Weather data is one of the main inputs in building energy simulation, and it can significantly influence the building energy balance. A comprehensive yearly analysis of a building's energy use requires the integration of hourly weather data. The approach is commonly used to analyze the energy efficiency of buildings or cities [88], [89], where 8,760 sets of hourly weather data are used for the entire year. However, this poses challenges when incorporating yearly meteorological variations into ML models. The primary challenge lies in combining the variable meteorological data with building features for the preparation of training and testing datasets in predictive model development. Due to the huge size of datasets when considering a full variation of meteorological data combined with building features using the artificially generated physics-based building energy result. For instance, when considering six meteorological parameters and ten building energy features, a single building's machine learning model generates over 82 million data points for year-round 8760 hourly data points (as discussed in Appendix B). Conducting hourly analyses for an entire year requires significant computational resources. It may not be feasible or necessary to analyze all hours in the whole year. Therefore, instead of a complex day-to-day building energy simulation (BES), we recommend simulations for several representative days.

Weather clustering was used to identify representative days capturing outdoor weather variables such as air temperature, solar radiation, wind speed, and relative humidity (RH) for the entire year. The weather clustering techniques have been widely used in various fields, including air quality trend studies [90], and operation strategies optimization [91], [92]. And weather pattern analysis [93], [94], [95], [96], [97]. Souayfane et al. [91] applied weather clustering with an optimization approach to control HVAC systems. Lundell et al. [98] identified six weather regions within the United States based on weather characteristics and analyzed patterns in forecast accuracy. Babanov et al. [99] compared four commonly used clustering methods to analyze weather regimes in the Euro-Atlantic region. Klampanos et al. [93] utilized *k*-means clustering to derive weather patterns.

Hoffmann and Schlünzen [94] employed various clustering methods to represent Urban Heat Island (UHI) patterns in present and future climates. Onal et al. [95] applied the *k*-means clustering method to examine a general pattern of weather data using approximately 8000 weather sensors in the United States for a single day. Luo [100] proposed a novel clustering-enhanced adaptive ANN model for day-ahead cooling energy forecasting, using *k*-means clustering to identify daily weather patterns and categorize annual datasets into clusters, with each cluster used to train an ANN sub-model. In the meteorological field, there is no one best method for the classification of weather patterns. The *k*-means method, however, ranks highly in these studies [94]. Peng et al. [101] employed cluster analysis algorithms to examine climate datasets from 270 cities, aiming to understand how building heating and cooling energy demand intensity is distributed geographically in relation to climatic characteristics.

Developing ML-based building energy prediction models on an hourly basis is challenging due to the large datasets required. A great deal of synthetic data will be generated by combining long-term hourly meteorological data with building characteristics under uncertainty. For instance, considering six meteorological variables and ten building features for a single building model generates over 82 million data points based on year-round hourly data (8,760 hours). Such vast datasets necessitate efficient data handling, posing a considerable computational burden. Many previous studies have relied on full-year datasets, which exacerbate these challenges. For instance, Gao et al. [102] utilized 81 million data points simulated for 16 commercial building types across 936 cities, considering 8,760 hourly data points per building for a full year. Using full-year data often results in redundant information, computational inefficiency, and scalability issues, especially for large-scale urban studies or when exploring multiple scenarios. To overcome this challenge, we introduced a methodology utilizing weather clustering techniques to identify representative days from long-term historical meteorological data on an hourly basis. The size of the final dataset was significantly reduced compared to day-by-day simulations. We streamlined the dataset while maintaining the original data integrity and variability. While weather clustering has been used in various fields like air quality studies and operational optimization, our solution marks the first application of weather clustering methods specifically for developing ML models in the context of building energy consumption. This innovative approach provides a generalizable framework that other researchers and engineers can replicate in different urban areas, aiding in the optimization of building design and retrofit phases.

4.2 Methodology

In this study, we introduce an innovative method applying weather clustering to develop ML models for predicting building energy consumption. The main methodology that we developed and implemented is shown in Figure 4-1. Three main steps were (a) the weather clustering methodology, (b) the generation of simulated data based on a physics-based building energy model, and (c) the development of an ML building energy prediction model. The methodology will be applied in typical residential buildings located in Doha, Qatar.

This work applied a k-means-based weather clustering technique to find representative days capturing the variability of yearly weather features from historical data. As shown in Figure 4-2, similar items are grouped into clusters by minimizing the squared distance between each object and its cluster mean value.

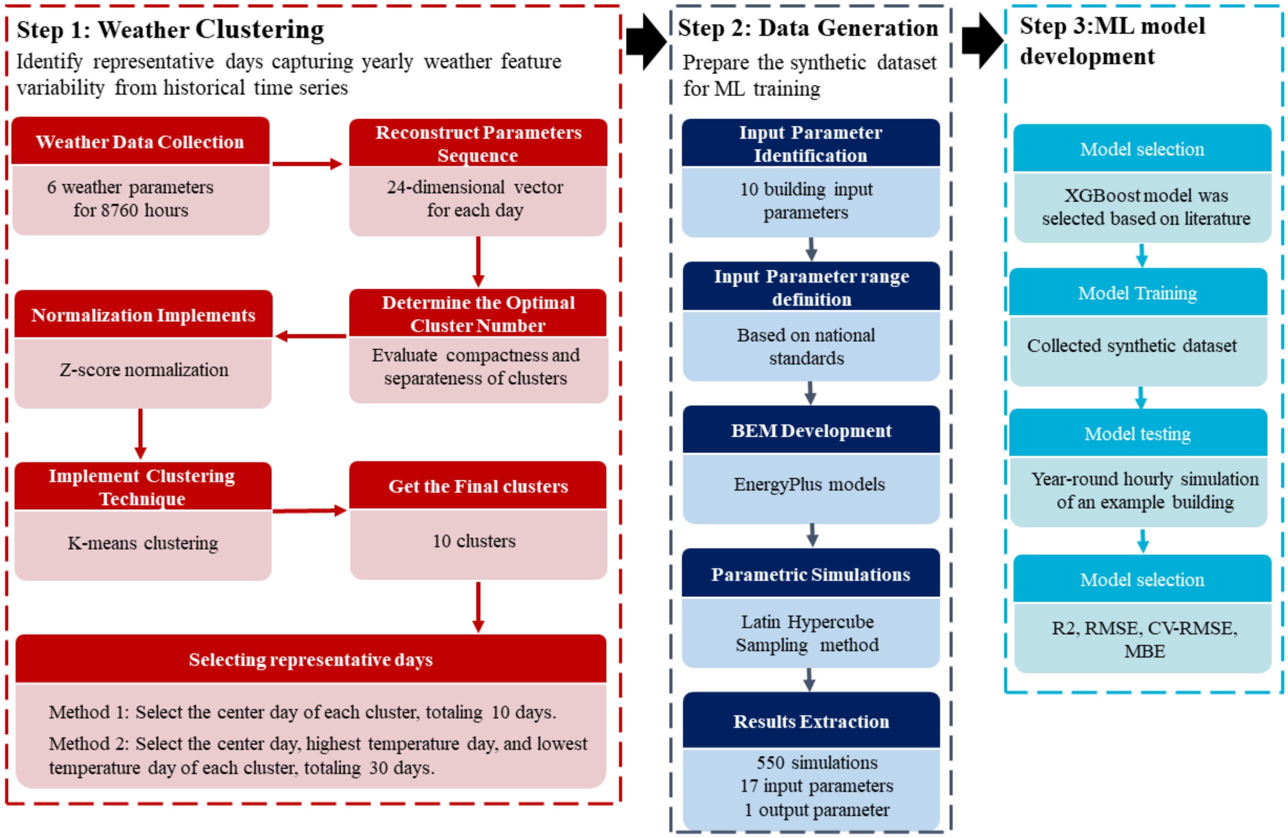


Figure 4-1. Flow chart of the model framework.

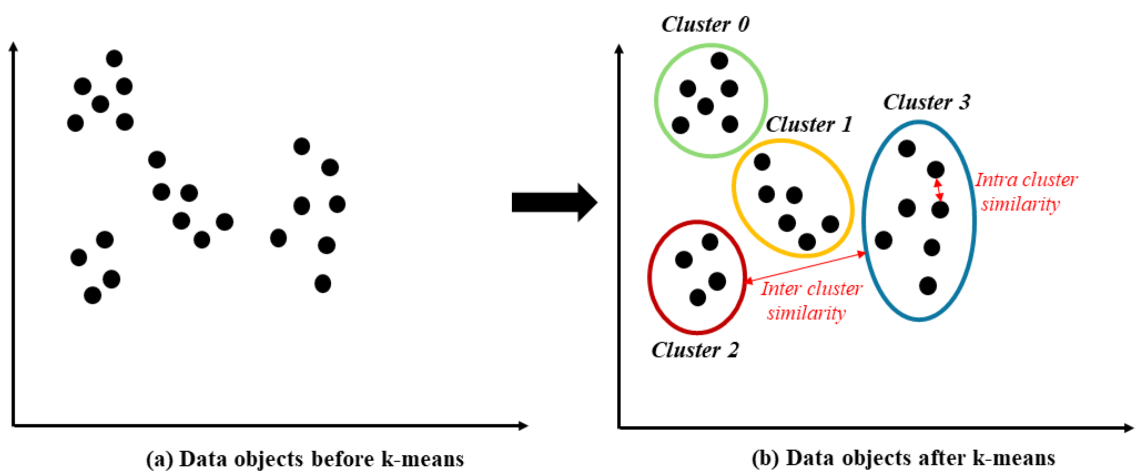


Figure 4-2. K-means clustering algorithm.

4.2.1 Data collection

The weather conditions observation has been conducted by the researchers for a variety of purposes, as weather serves as a fundamental input for numerous simulations and the study of weather phenomena itself. In the context of developing machine learning building energy models at the design phase, the EnergyPlus weather data (EPW) was used. This data represents compiled long-term observations collected at the Doha International Airport meteorological station. The typical meteorological year (TMY) weather data is used since it is derived from long-term measurements and is generally used to design building energy efficiency systems. Table 4-1 lists the weather parameters that were selected for clustering the annual meteorological data using the statistical clustering *k*-means method.

A residential building in Doha, Qatar serves as the case study. All building elements and their ranges that are included in the EnergyPlus model are summarized in Table 4-1. The information was collected from a variety of sources, such as relevant codes and standards, current building practices, and prior literature. All these parameters are collected from local standards [69], [74], ASHRAE standards [72], [103], and literature references [104], [105], [106].

Table 4-1. Building energy model's inputs and range.

	Parameter	Unit	Range
Weather profile	Outdoor air dry-bulb temperature	°C	(9, 46)
	Outdoor air relative humidity	%	(5, 99)
	Wind speed	m/s	(0, 12)
	Wind direction	°	(0, 354)
	Normal solar irradiation	Wh/m^2	(0, 474)
	Diffuse solar irradiation	Wh/m^2	(0, 681)
Building envelope characteristics	Wall U value [26-29]	W/m^{2}K	(0.3, 0.7)
	Roof U value[74] [72]	W/m^{2}K	(0.19, 0.44)
	Window U value[69] [72]	W/m^{2}K	(1.8, 2.8)

	Window SHGC[69] [72]	/	(0.25, 0.275)
Internal loads	Occupancy density [72] [104]	Person/100m^2	(2.83, 76.9)
	Appliance power density[72] [104]	W/m^2	(2, 8)
	Lighting power density[69], [72] [104]	W/m^2	(0, 6.5)
HVAC settings	Ventilation rate [72], [103] [104]	cfm/ft^2	(0.06, 0.098)
	Heating set point[69] [104], [105], [106]	°C	(24, 28)
	Cooling set point[69] [104], [105], [106]	°C	(25, 28)

4.2.2 Weather clustering techniques

While weather patterns can differ significantly from place to place, for a given location and season, they share similar characteristics [92]. There is the possibility of extracting representative days from the entire year while still covering yearly weather patterns. To achieve this, a clustering analysis is employed as an unsupervised data mining approach. Several representative days should be selected from the entire year. Our approach employed *k*-means clustering since it is a popular and simple method of identifying groups of data points that have similar characteristics. While *k*-means is known to be sensitive to noise and outliers, our application deviates from conventional clustering scenarios in that we cannot disregard data that is rarely observed as noise. Instead, we must take care to manage extreme weather conditions, despite their relative rarity. Extreme weather events can be evaluated to examine their impact on city energy consumption, agriculture productivity, transportation systems, and human thermal comfort. Consequently, we propose a weather clustering procedure, that can handle all scenarios, which aims to select a representative sample of days to represent the entire year, including several weather parameters, especially the weather clustering technique in the development of machine learning buildings energy prediction models. The process of weather clustering is as follows.

1) Reconstruct weather parameter sequences.

2) Normalization implements.

3) Determine the optimal clustering number.

4) Implement *k*-means clustering.

Step 1: Convert weather parameters form

As aforementioned, the TMY weather in Qatar was used as a weather database on an hourly basis. A diurnal cycle is reconstructed for each climate variable. As a result, all the weather parameters are converted into a 24-dimensional vector for each day, all the climate parameters are listed in sequence in terms of air temperature, RH, wind speed, wind direction, normal solar irradiance, and diffuse irradiance.

Step 2: Normalization implementations

Normalization is used to eliminate the impact of dimensions because several features often have different dimensions. All parameters listed in Table 4-1 were normalized using Z-Score normalization. In comparison with min-max normalization, this method is less affected by outliers. Since it scales data based on the mean and standard deviation, as demonstrated in Equation 4-1, outliers have less impact on the normalized values. Using Jan. 1 as an example day, Figure 4-3 shows the normalized attribute values.

$$z = \frac{x - \mu}{\sigma} \qquad \text{Equation 4-1}$$

where x represents the unstandardized weather parameters, μ is the mean value of each actual weather parameter, σ is the standard deviation of each parameter, and z represents the standardized data. When a value is exactly equal to the mean of all values of a feature, it will be normalized to 0. A number below the mean will be negative; one above the mean will be positive.

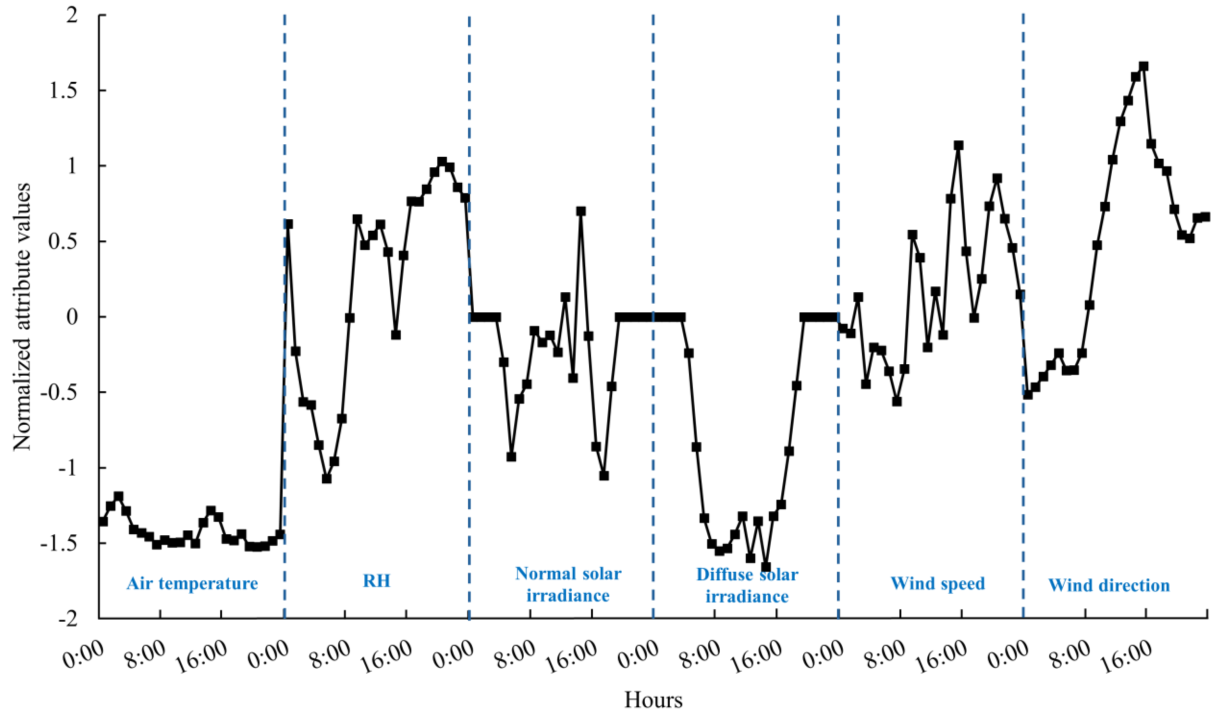


Figure 4-3. Normalized weather profile of Jan.1.

Step 3: Determine the optimal number of clusters

Determining an optimal number of k has to be prescribed for the k-means method. An algorithm should be used to calculate the number of clusters k based on the data set and the criteria [107]. Generally, it is preferred to have a small number of clusters with examples scattered around them in a balanced way [108]. A small value of k may aggregate many natural clusters that hide desirable features, causing genes with different expression patterns to end up in the same cluster. Therefore, classifications with small k might be considered circulations or weather regimes rather than weather patterns [94]. On the other hand, a large value of k will lead to many trivial clusters with similarly expressed genes being placed in different clusters. In either case, the clustering results do not result in optimal detection of all interesting features [109], [110].

Previous studies have employed cluster validity indices (CVIs) to quantify the ideal number of clusters. It is expected that clusters will be modest in diameter and far apart from one another. The validity index is defined as the ratio of these two distances [29]. Plotting the CVIs versus the cluster number yields the ideal number of clusters to have when the CVIs reach their minimal value. An illustration of the dynamic validity index (DVIndex) may be found in [29]:

$$DVIndex = \min_{k=4,\ldots,K}\{IntraRatio\,(k) + \gamma InterRatio(k)\} \quad \text{Equation 4-2}$$

$$IntraRatio(k) = \frac{Intra(k)}{MaxIntra} \quad \text{Equation 4-3}$$

$$InterRatio(k) = \frac{Inter(k)}{MaxInter} \quad \text{Equation 4-4}$$

where *k* is the cluster number, starting from 4 in this work, and z_i denotes the center of the cluster C_i. $i = 1, 2, \ldots, k-1$; $j = i+1, i+2, \ldots, k$.

IntraRatio and *InterRatio* represent the overall cluster's compactness and separateness respectively. The normalized ratios are used for comparison, ranging from 0 to 1. The detailed equations for the computation of each component mentioned above are provided in the Appendix A section for reference.

Step 4: Implement k-means clustering

k-means clustering finds the most representative data point in the cluster, including six main steps: 1) select k random centroids; 2) assign each data point to the closest centroid; 3) re-calculate centroid of each cluster; 4) repeat steps 2&3; 5) terminate when converged. The process iterates until re-calculation does not change the centroid of clusters. The centroids are the best representation of their clusters. In the k-means clustering method, centroids are the median of all data points in a cluster. The *k*-means method is calculated here in RapdMiner Studio software, which is one of the best data mining tools.

4.2.3 Generation of synthetic data

We are aiming to develop a building energy prediction model that predates the construction phase itself so that planners can manage to estimate the building energy consumption. In this context, the energy data required in ML model development is inaccessible in building design processes. Thus, this study utilizes artificial data for the training and testing of ML models.

The EnergyPlus model was developed and serves as a detailed BES model. It performs a warm-up on the first day of the simulation period to achieve thermal stability, improve accuracy, and ensure consistent and reliable results. The building hourly EUI (energy use intensity) was selected as the target output in this work. Based on the impact of input variables on the outputs, input variables can be determined. Considering too many input variables can increase computational time and costs, while too few or irrelevant ones can limit the machine learning model's capabilities and accuracy [111]. A BES is typically made of tens or hundreds of attributes influencing its energy consumption, in this study, the input variables are collected based on the sensitivity analysis results in published results. Building energy models (BEM) involve many input parameters influencing energy consumption, including weather conditions, building geometry and properties, building system, and occupant information [112], [113], [114]. The current body of work focuses on inputs related to five main categories: building envelopes, internal loads, schedules, HAVC systems, and outdoor weather. Synthetic datasets were generated using the methodology described by Jia et al [56], creating a baseline model, and parametric simulations based on Latin Hypercube Sampling (LHS) method. The dataset generated in parametric simulations was divided into training and testing data. The training data were normalized and then imported to train the ML models, and testing data were used to evaluate the model's performance.

4.2.4 Hourly Machine-Learning building energy prediction model development

ML models are conventionally developed using a series of standard steps. It involves data preparation and cleaning, splitting the database into training and testing sets, training the ML model, performance evaluation, and finally the validation and deployment of the model. In our study, we are going to run XGBoost since its good accuracy and low computational cost reported in the literature [115], [116], [117].

By utilizing the second-order Taylor expansion of the loss function and incorporating a regularization term, XGBoost to find the optimal solution. This method balances the decrease of the objective function and the complexity of the model to prevent overfitting. The XGBoost model is described as follows:

$$\hat{y}_i = \sum_{k=1}^{K} f_k(x_i), f_k \in F \qquad \text{Equation 4-5}$$

where k is the number of decision trees, $f_k(x_i)$ is the function of input in the k-th decision tree, $\hat{y}_i$ is the predicted value.

The addition of each tree effectively introduces a new function $f_k(X, \theta_k)$ to fit the residual of the last prediction. After training with *K* trees, each feature in the prediction samples is assigned a leaf node in each tree, and each leaf node has an associated score. The final prediction value is obtained by aggregating the scores from all the trees. The flow chart of XGBoost is depicted in Figure 4-4.

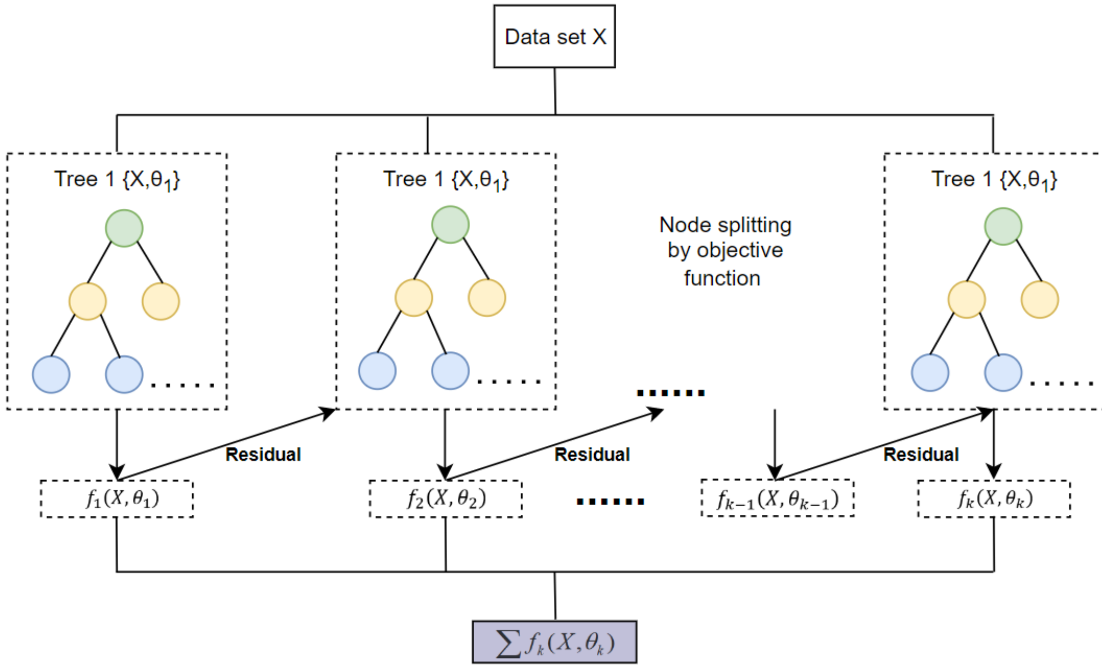


Figure 4-4. Flow chart of XGBoost.

To evaluate the model's performance, four statistical indices were employed, the mean bias error (MBE) (Equation 4-6), the coefficient of determination (R^2) (Equation 4-7), root mean square error (RMSE) (Equation 4-8), and coefficient of variation of RMSE [CV(RMSE)] (Equation 4-9).

$$MBE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i) \times 100(\%) \qquad \text{Equation 4-6}$$

$$R^2 = 1 - \frac{\Sigma_{i=1}^{n}(\hat{y}_i - \bar{y})^2}{\Sigma_{i=1}^{n}(y_i - \bar{y})^2} \qquad \text{Equation 4-7}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}} \qquad \text{Equation 4-8}$$

$$\text{CV(RMSE)} = \frac{1}{\bar{y}}\sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}} \times 100(\%) \qquad \text{Equation 4-9}$$

where $\hat{y}_i$ denotes predicted variable value for period i, y_i represents the observed value for period

i. $\bar{y}$ is the mean value of the observed value. n is the sample size.

MBE and CV(RMSE) were suggested by ASHRAE guidelines [118]. The MBE evaluates how accurately the model's predicted energy use aligns with the actual metered data on a monthly or annual basis. However, since MBE can be affected by offsetting errors, an additional metric is often needed. The CV(RMSE) is obtained by dividing the root mean square error (RMSE) by the mean of the measured data. This metric assesses the model's fit to the data, with a lower CV(RMSE) indicating a better model performance. In general, it is easier to achieve a low MBE than a low CV(RMSE). Typically, models are considered calibrated if they produce MBEs within ±10% and CV(RMSE)s within ±30% when using hourly data [118].

4.3 Results and discussion

To evaluate the accuracy and effectiveness of the proposed weather clustering-based hourly building energy prediction model, the clustering results of the daily weather data profile are examined in detail. Then, the performance of the XGBoost model was investigated. Finally, a developed model is applied to predict your-round building performance of one building, and the predicted cooling demands of the example days in each season are evaluated. The simulation relies on Qatar's TMY weather data. This study is conducted using Jupyter and RapidMiner, and it is executed on a computer equipped with an Intel(R) Core (TM) i7-10750H CPU @ 2.60GHz and 16GB of memory, operating on Windows 10 Pro.

4.3.1 Weather clustering results

In this study, we integrate this DVIndex into the traditional k-means clustering algorithm to find out the representative days out of the entire year. We define the k varies from 4 to 20. Figure 4-5 illustrates the variation of DVIndex across cluster numbers ranging from 4 to 20, using a 365-day annual dataset. The clustering with the smallest DVIndex is considered the best one. Meanwhile, employing a small number of clusters (*k*) tends to represent general circulation or weather regimes rather than distinct weather patterns [109]. Hoffmann and Schlünzen decided that a cluster number of 6 or larger is accepted for k value, the cluster number of weather patterns [94]. By analyzing estimated cooling costs across various cluster numbers, a threshold of *k*=10 is deemed reasonable, as demonstrated by Saudi Arabia weather data [91]. The analysis reveals that the lowest DVIndex values are observed at cluster numbers 4, 5, and 6, followed by a comparable value at 10 clusters. Following the same restriction in the previous literature, after considering the trade-off between DVIndex and the number of clusters, the optimal choice for this scenario appears to be 10 clusters. The year-round database is divided into 10 clusters [94].

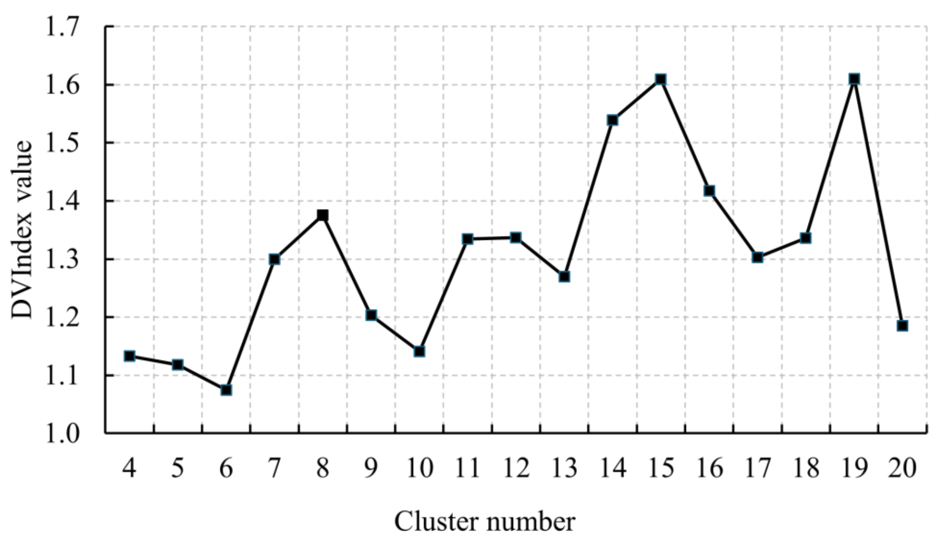


Figure 4-5. DVIndex variation in clustering procedure.

We applied t-distributed stochastic neighbor embedding (t-SNE) to visualize the outcomes of *k*-means clustering for the overall distribution of data points and the similarity maps As a result of t-SNE, high-dimensional data is projected onto low-dimensional space, preserving the distance between data points in a high-precision manner. As shown in Figure 4-6, the application of *k*-means resulted in the segmentation of data points into 10 distinct categories, whose representation in a two-dimensional space allows for a comprehensive visualization. Notably, the t-SNE plot elucidates the distinct nature of the 10 clusters, portraying them as non-overlapping entities.

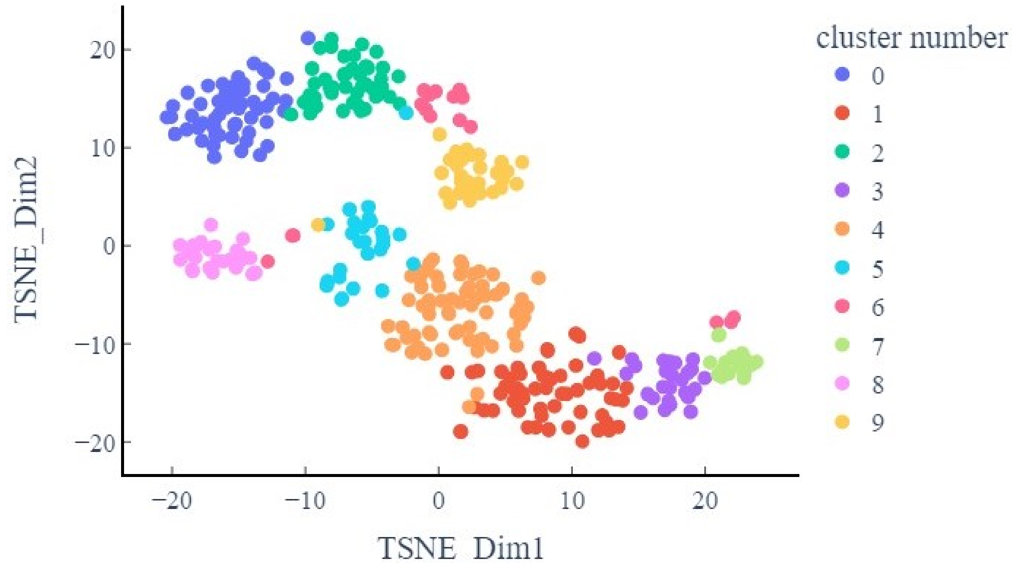
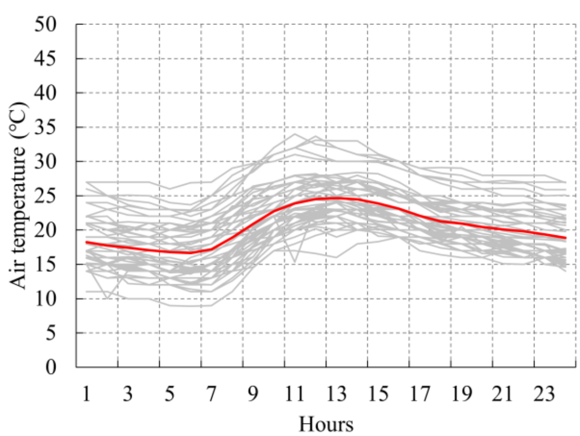
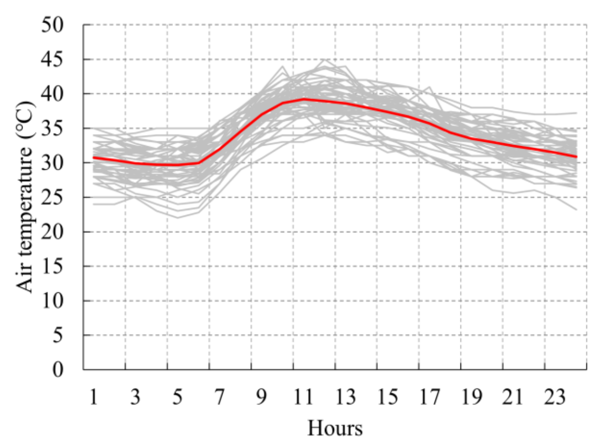


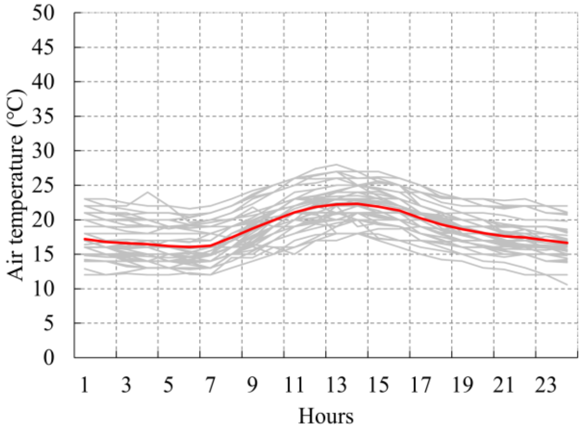
Figure 4-6. The t-SNE visualization of the *k*-means clustering results.



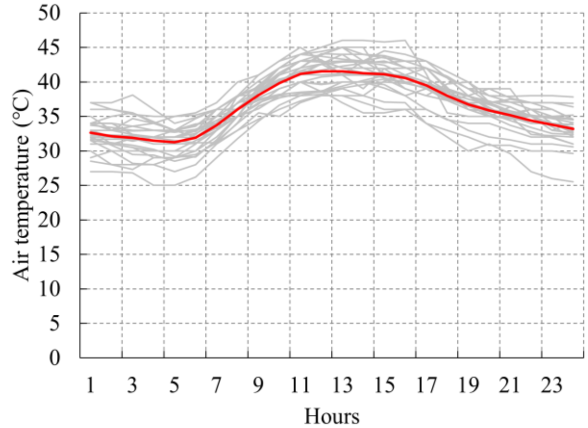
Cluster 0 (with 52 days)



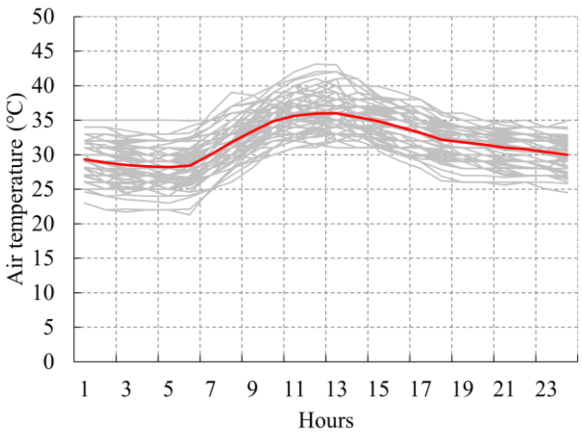
Cluster 1 (with 60 days)



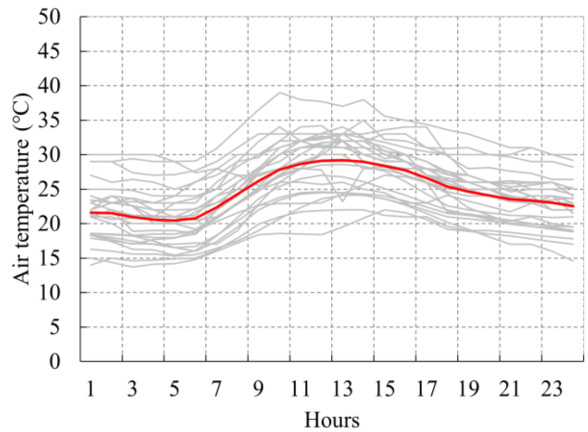
Cluster 2 (with 43 days)



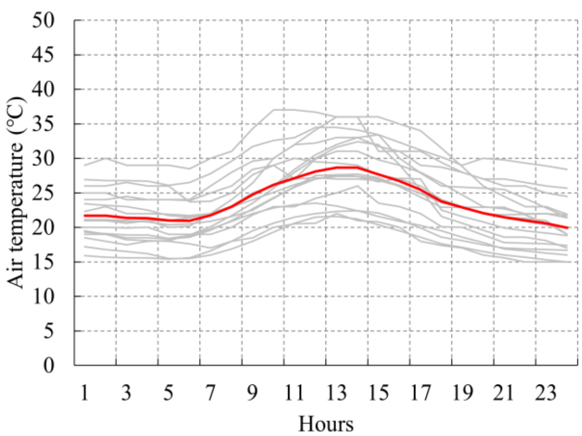
Cluster 3 (with 26 days)



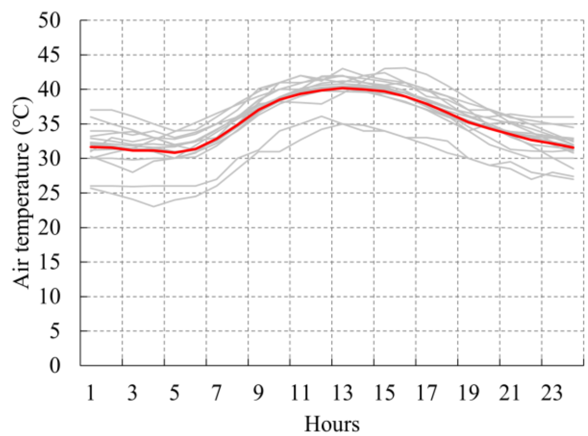
Cluster 4 with 68 days)



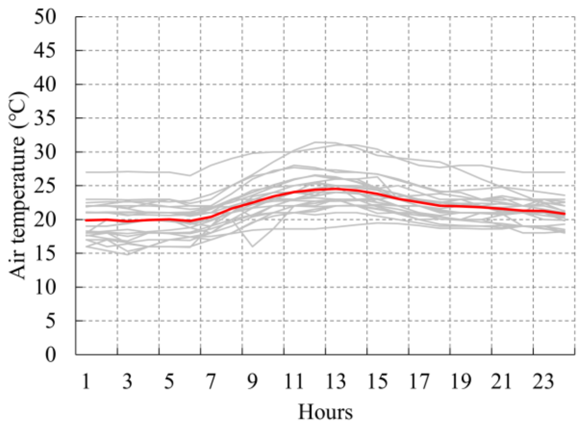
Cluster 5 (with 26 days)



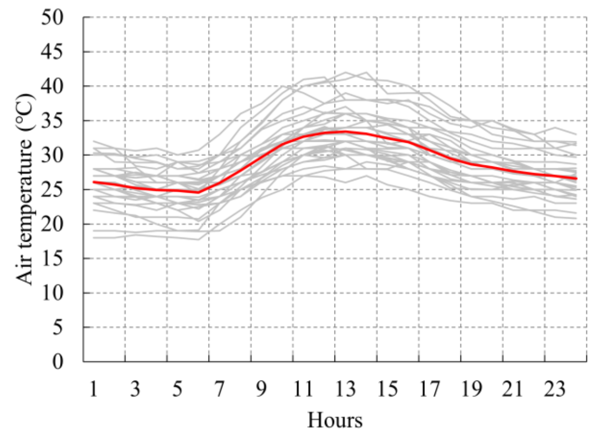
Cluster 6 (with 17 days)



Cluster 7 (with 17 days)



Cluster 8 (with 25 days)

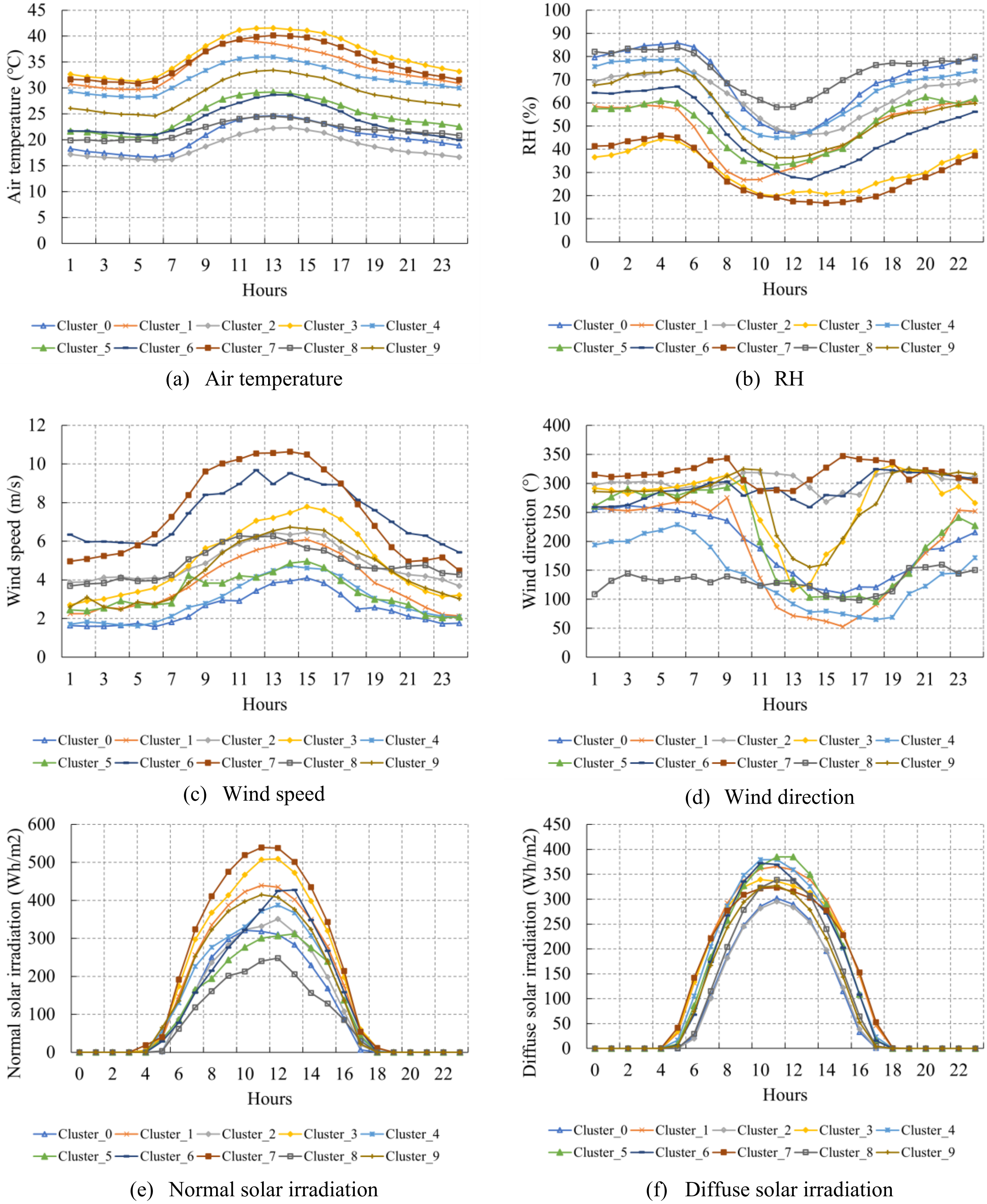


Cluster 9 (with 31 days)

Figure 4-7. Clustering results for the air temperature data attribute.

The year-round Doha's TMY database was divided into 10 groups using a value of k-10. Each group is represented by a real-period representative day chosen from the cluster genes. Cluster 4 has 68 days in total, which is the greatest number of days. In contrast, the fewest number of days (17 days) are found in clusters 6 and 7. Figure 4-7 depicts the k-means clustering for the air temperature portion, the representative day is shown by the red line, and other genes are represented by grey curves. The cluster center is determined by identifying the data point with the shortest distance to all other points within the cluster. In this study, the cluster center is defined as the center day of each cluster. Taking air temperature as an example, the outdoor air dry-bulb temperature decreases slightly during the first 7th hour, then increases till the 13th hour, and finally decreases till the end of the day.

The characteristics of these center days are shown in Figure 4-8. For instance, Cluster_3 is characterized by high air temperature, low relative humidity, and high solar irradiation, while Cluster_0 has low air temperature, high humidity, low wind speed, and low diffuse solar irradiation. Cluster_7 features high air temperature, low humidity, high wind speed, high wind direction angle, and high solar irradiation.



(a) Air temperature (b) RH

(c) Wind speed (d) Wind direction

(e) Normal solar irradiation (f) Diffuse solar irradiation

Figure 4-8. Hourly weather data profiles for the ten cluster centroids

4.3.2 Synthetic data generation

Following the identification of representative days, the synthetic database was generated to build

ML models. The LHS method was applied to produce random samples of input parameter combinations within specified ranges for various parameters. We developed an R Script to automate the process of EnergyPlus model generation, input sample generation, EnergyPlus simulation, and data extraction. All building elements and their ranges that are included in the EnergyPlus model are summarized in Table 1. A total of 550 EnergyPlus simulations were conducted.

To determine the optimal training dataset for an XGBoost building energy consumption prediction model, two methods of dataset preparation were implemented and compared. Method 1 involved conducting EnergyPlus hourly simulations for the center day of each cluster, totaling 10 days and comprising 2,376,000 data points. Method 2 involved performing EnergyPlus hourly simulations for the center days, as well as the days with the maximum and minimum temperatures in each cluster, totaling 30 days and comprising 7,128,000 data points. To test whether the developed weather clustering-based XGBoost model could predict year-round hourly building cooling demand, one testing case was selected out of 550 simulations, resulting in 8,760 hourly simulations. To ensure the model was tested on unseen data, the 30 days used in Method 2 were removed, resulting in a final testing dataset of 335 days, or 8,040 hours of data. Table 4-2 summarizes the details regarding the training and testing datasets.

Table 4-2. Input and output datasets for training and testing.

	Parameters	Training	Testing
Input datasets	Weather profiles	Method 1 & 2	TMY weather profile from Doha International Airport weather station
	Building envelope characteristics	Local and international standards, and existing literature	
	Internal loads		
	HVAC settings		
	Hour of the day	Simulation calendar	

Output datasets	Simulated cooling demand	Simulated results from EnergyPlus

4.3.3 Building energy prediction model development

XGBoost model was developed using the data generated according to Section 3.2. 70% of the generated synthetic data is used for training the ML models. The remaining 30% is applied for model testing. During the testing phase, the developed prediction function and inputs are used to predict hourly EUI outputs. We start by selecting the centroid day of each cluster to represent the weather boundary conditions. Then, within each cluster, we choose three days: the centroid day, the day with the minimum air dry-bulb temperature, and the day with the maximum dry-bulb temperature.

This step aims to assess the predictive capability of the developed model for building energy consumption. Only machine learning models that closely match outputs from detailed computer models within an acceptable error range, using the same inputs, can potentially replace detailed computer models in future analyses for their intended purposes. For testing purposes, a single building was randomly selected from the 550 combinations of building features generated during the parametric simulation phase. In Method 1, simulations used the center day of each cluster as weather conditions in BEM simulations, with each case running for 10 days. The machine learning model trained using Method 1 achieved an R2 value of 0.85 and a root mean square error (RMSE) of 0.006 kWh/m2. Additionally, the mean bias error (MBE) was found to be -1.41%, and the CV(RMSE) was 31.19 (as shown in Table 4-3). Method 2, which incorporated the center day, maximum temperature day, and minimum temperature day as boundary conditions for BEM simulations, showed improvements in all error metrics. The machine learning model achieved an R2 value of 0.94 and an RMSE of 0.004 kWh/m2. The MBE and CV(RMSE) values were -1.66% and 19.31%, respectively, which are considered reasonable given the ±10% tolerance for MBE and ±30% for CV(RMSE) as required by ASHRAE Guideline 14 [118]. MBE represents the average deviation of all predictions across the samples. A positive MBE indicates the model tends to overpredict, while a negative MBE indicates underprediction. The closer MBE is to zero, the smaller the deviation between the model's predictions and the actual values. Compared to existing studies, our model shows competitive results. Wang et al. [119] reported R^2 values around 0.912 and CV-RMSE near 20% for synthetic buildings energy uses prediction model, with slightly lower performance for existing buildings ($R^2 \approx 0.83$, CV-RMSE ≈ 21.65%). Hong et al. [120] achieved a mean CV-RMSE of 24.15% using a K-nearest neighbor (KNN) algorithm to predict hourly energy consumption. Dong et al. [121] demonstrated CV-RMSE values decreasing from 24.1% to 19.8% as data availability increased from 20% to 100%. In comparison, our proposed model achieved an R^2 value of 0.94 and a CV-RMSE of 19.31%, outperforming the reported models in terms of accuracy and predictive reliability.

Table 4-3. Performance of developed ML model.

	R2	RMSE	CV-RMSE	MBE
Method 1	0.85	0.006	31.19%	-1.41%
Method 2	0.94	0.004	19.31%	-1.66%
ASHRAE Guideline tolerance	/	/	30%	±10%

Figure 4-9 (a) and (c) compare the predicted value generated using Method 1 and Method 2 against simulation results from EnergyPlus in the testing dataset. The red dashed line represents the ideal case (where predicted and simulated values perfectly match). The green and orange lines represent upper and lower 15% error margins. Method 2's predictions are more closely aligned with the ideal line and fall within the upper and lower 15% error margins. While there is a noticeable dispersion from Methos 1, especially at higher EUI values. According to Figure 4-9 (b) and (d), the error distribution for Method 1 is highly centralized around 0, indicating minimal deviation from the

EnergyPlus simulation results, but with a slight underestimation. The comparative analysis highlights that Method 2 achieves higher accuracy and consistency in its predictions compared to Method 1.

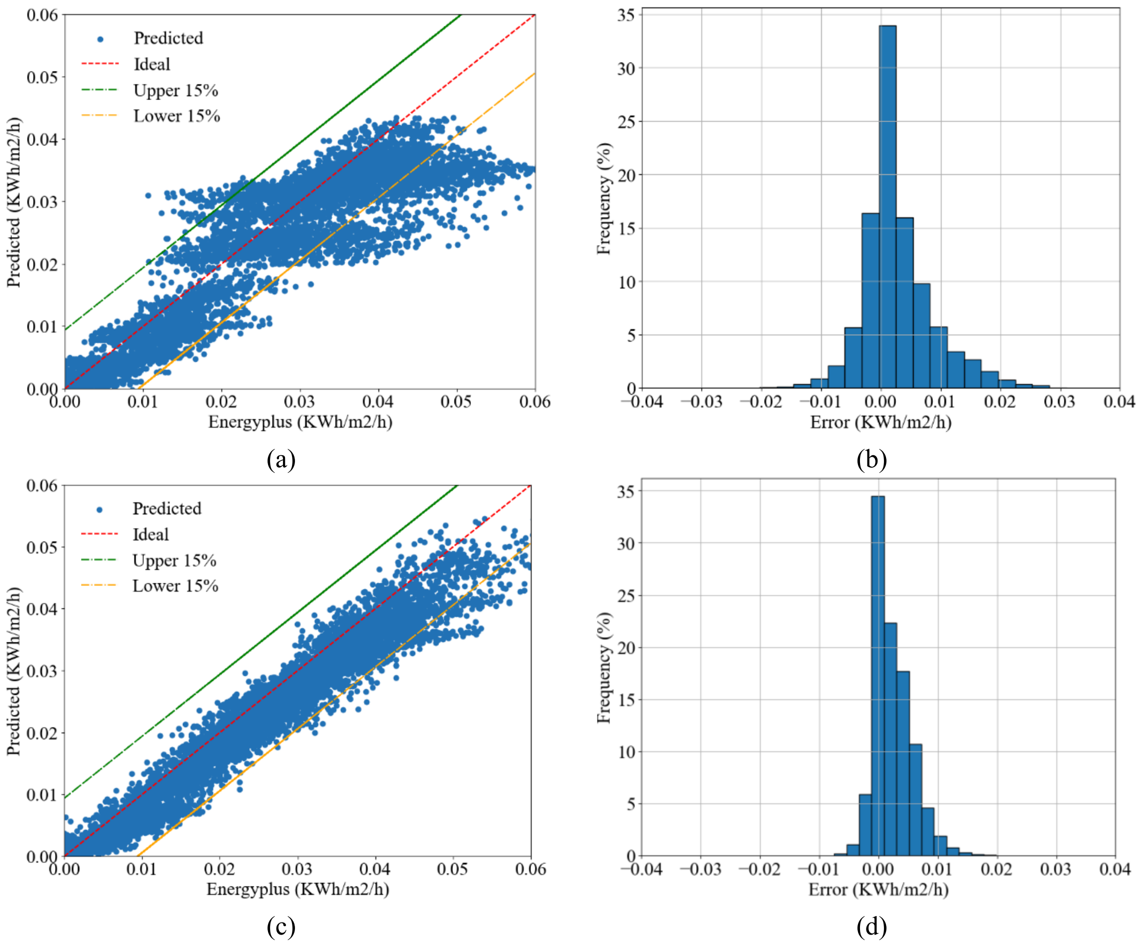
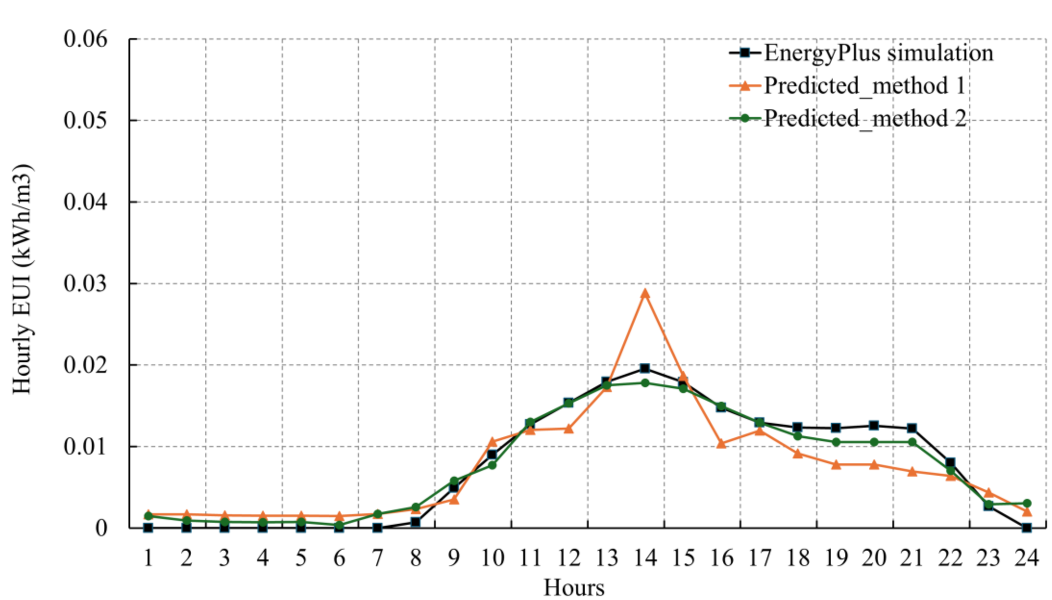
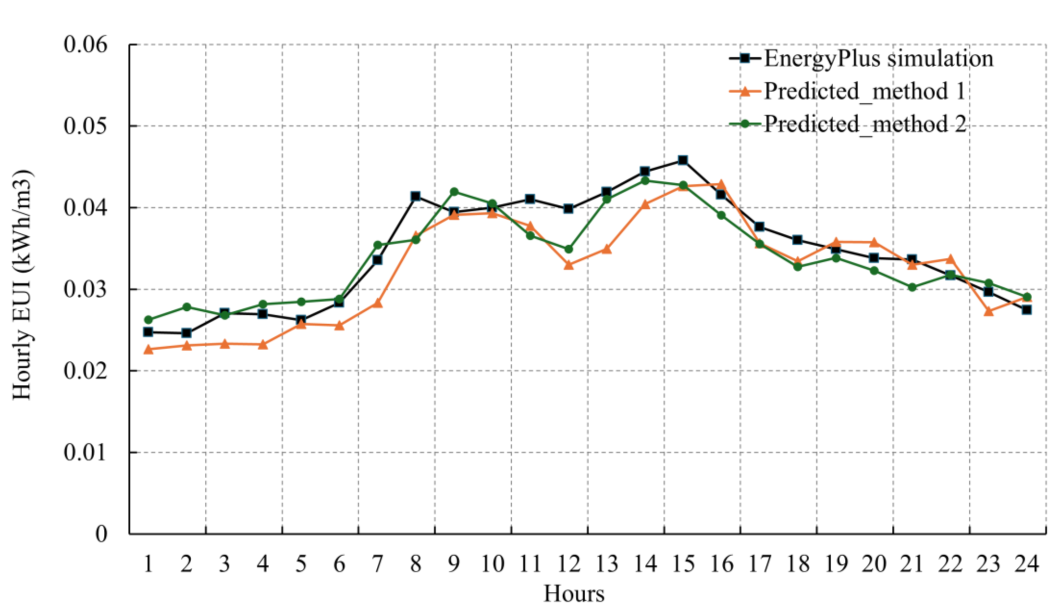


Figure 4-9. Comparison of Actual vs. Predicted Building Cooling Energy Demand and Error Distribution. (a) Actual and predicted building cooling energy demand using Method 1; (b) Error histogram for Method 1; (c) Actual and predicted building cooling energy demand using Method 2; (d) Error histogram for Method 2.

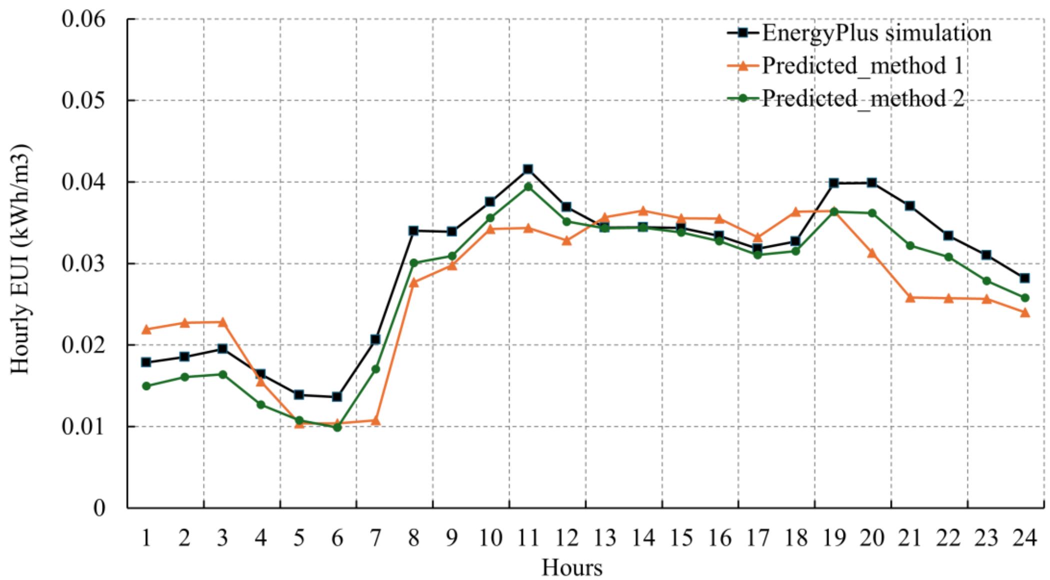
It is challenging to anticipate peak consumption with accuracy, even with advances in machine learning techniques [122]. However, the capacity to forecast peak demand might help energy management systems implement the best plans of action during demand response incidents. To take a deep look into the hourly building cooling energy performance, the cooling demand prediction results from 4 example days in each season are presented in Figure 4-10. Qatar's weather is divided into four seasons: summer (June to August), autumn (September to October), winter (November to February), and spring (March to May). The chosen example days are, August 16th for summer, September 17th for autumn, December 18th for winter, and March 21st for spring. Method 2 consistently provides more accurate and stable predictions across all seasons, closely matching the EnergyPlus simulations. Method 1 shows reasonable accuracy but has noticeable discrepancies during the peak hours.



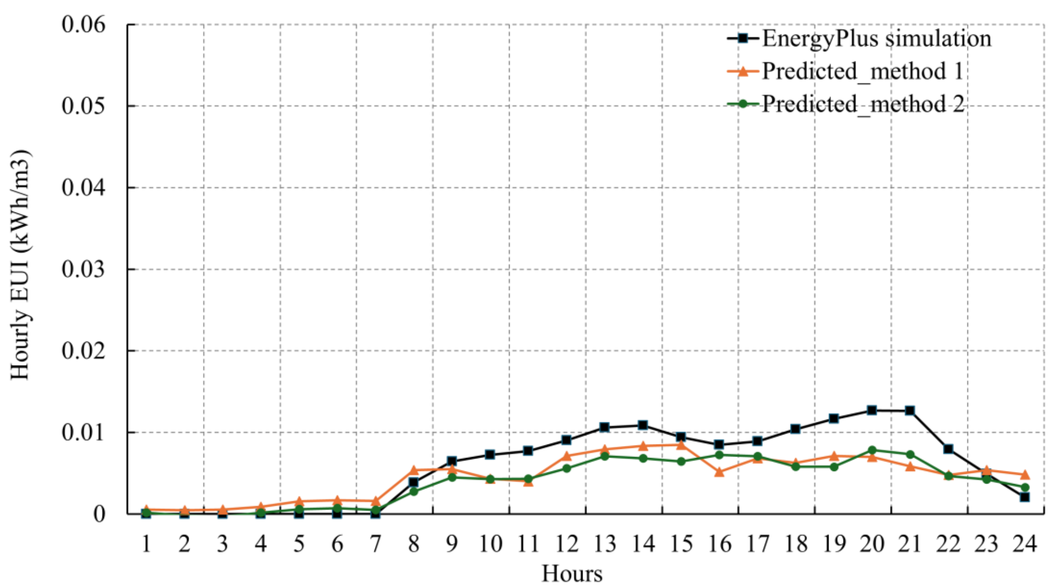
(a) Spring



(b) Summer



(c) Autumn



(d) Winter

Figure 4-10. Prediction from machine learning models and simulated values from EnergyPlus models for cooling energy.

Weather clustering techniques are applied to identify representative days that capture yearly weather variations, enabling a reduction in the need for computationally intensive day-by-day simulations. This approach integrates with the capabilities of advanced digital technologies such as building information modeling (BIM) [123] and digital twin systems [124], supporting the ongoing trend toward digitalization in the construction and building management sectors. By integrating weather clustering within these frameworks, BIM and digital twins can deliver more resource-efficient insights during the design phase. This is especially advantageous for implementing and optimizing emerging technologies, like photovoltaic (PV) panels [125], [126] and wind turbines [127], within building systems. For example, through weather clustering, BIM and digital twins can focus analyses on representative days with varying solar radiation levels to analyze PV performance, or on days with different wind speeds and directions to estimate wind turbine efficiency. Furthermore, weather clustering can aid in identifying and managing extreme weather scenarios, contributing to the resilience planning and energy efficiency assessments crucial for meeting sustainability and carbon neutrality goals.

4.4 Summary

Hourly building energy prediction is essential to efficient building energy management, aiding tasks like predicting peak loads, comparing energy systems, and optimization during the design phase. This paper explores the advantages of ML approaches for predicting hourly building energy consumption during the design phase. We applied weather clustering techniques to reduce the vast dataset of synthetic data combining long-term hourly meteorological data with building characteristics.

Synthetic data is commonly used for training and testing ML models when real-life data is unavailable, but it presents challenges due to the vast datasets generated from long-term hourly meteorological data combined with building characteristics. Our method applies weather clustering techniques to identify a subset of representative days from yearly weather data, thereby mitigating computational complexity by focusing computational efforts on a reduced set of days. Subsequently, synthetic data is generated based on physics-based BEMs under selected representative days, which are then utilized for the training and evaluation of an ML model. A comprehensive dataset comprising 6 meteorological parameters, 10 building features, and hourly building cooling energy consumption is compiled, amounting to a total of data points. Ten weather patterns were identified from long-term hourly meteorological data. To determine the optimal training dataset for an XGBoost building energy consumption prediction model, two methods were compared. Method 1 used EnergyPlus hourly simulations for the center day of each cluster (10 days, 2,376,000 data points). Method 2 included simulations for the center days, and the maximum and minimum temperature days in each cluster (30 days, 7,128,000 data points). We drastically reduce the dataset size from 82 million data points for the studied building, by selecting one/three days in each cluster. Employing the XGBoost model, peak load and hourly building cooling load are predicted in this study.

A residential building in Doha, Qatar serves as the case study. Throughout the testing phase, Method 2 outperformed Method 1. Method 2 considered not only the center day of each cluster but also the maximum and minimum temperature days of each cluster. Method 2 consistently provides more accurate and stable predictions across all seasons, closely matching the EnergyPlus simulations. Method 1 shows reasonable accuracy but has noticeable discrepancies, particularly in spring during the afternoon. The model using Method 2 is characterized by an MBE value of -1.66% and a CV-RMSE value of 19.31%. These metrics are within the acceptable thresholds of 10% and 30% in ASHRAE guideline 14, proving that the model is reliable and robust in forecasting building cooling energy consumption. Moreover, Method 2 provides more accurate predictions during the peak hours across all seasons than Method 1. The presented approach contributes by employing weather clustering to effectively downsize the database of ML-based hourly building energy prediction models while maintaining acceptable accuracy levels, thereby aiding engineers and city planners during design and retrofit phases. Further sensitivity analysis. This model can be used in place of BEM to predict building energy performance at a much lower computational cost.

Chapter 5. Assessment of urban microclimate at urban scale

This chapter attempts to evaluate the effects of urban microclimate on building ambient environment with a high degree of spatiotemporal granularity at a city scale. A novel integrated platform was employed to facilitate the information exchange between UBEM and UCM models and running simulations with CityFFD (City Fast Fluid Dynamics), an urban-scale fast fluid dynamics model for microclimate modeling, and CityBES (City Buildings, Energy, and Sustainability) for building energy model. A new flexible and tool-agnostic data schema (in JSON) is used to facilitate the exchange of data between urban building energy models and urban microclimate models. The entire city of San Francisco was examined as a case study, encompassing 148,698 buildings. The simulation was conducted during the hottest hour of September 1, 2017, a heatwave day with record-breaking temperatures. We investigated the impacts of urban microclimate on the ambient atmosphere of buildings[4].

5.1 Introduction

During building/city design and retrofit phases, it is crucial to estimate building energy consumption accurately to provide city planners and policymakers with detailed information on energy use. To this end, Urban Building Energy Modelling (UBEM) tools are widely used to estimate building energy consumption based on a bottom-up physics-based approach. UBEM can simulate each physical building in a district or a city's entire building stock while considering

[4] This chapter has included the contribution of the author in multiple publications:
1. Dongxue Zhan, Wanni Zhang, Maher Albettar, Mohammad Motezazadeh, Na Luo, Tianzhen Hong, Liangzhu (Leon) Wang (2024). "Integrated UBEM and UCM assessment to understand urban microclimate under heatwaves: A case study of city of San Francisco." The 1st International Conference of Net Zero Carbon Built Environment, Nottingham, UK, 3-5 July 2024.
2. Na Luo, Xuan Luo, Mohammad Mortezazadeh, Maher Albettar, Wanni Zhang, Dongxue Zhan, Liangzhu (Leon) Wang, and Tianzhen Hong (2022). "A Data Schema for Exchanging Information between Urban Building Energy Models and Urban Microclimate Models in Coupled Simulations." Journal of Building Performance Simulation, November, 1–18. doi:10.1080/19401493.2022.2142295.

urban microclimate impacts [128].

The urban microclimate is a small area around a building that has different atmospheric conditions than the surrounding area. Buildings in urban areas suffer from higher air temperatures due to the urban microclimate effect as well as reduced wind flow as a result of surrounding structures that block airflow. In addition, the reduced sky exposure and shaded solar heat by neighboring buildings alter the radiation balance of the urban environment [6]. These factors collectively affect the thermal and energy performance of urban buildings. A lack of knowledge of the local microclimate will decrease the accuracy of building energy simulation results. However, static Typical Meteorological Year (TMY) data have been widely used in BEM studies to represent the ambient climate of the building area without considering the local microclimate, omitting complex interactions between buildings and the environment [5]. From the perspective of UBEM, the urban microclimate could be obtained from local observational weather data, a data-driven forecasting model, and a physics-based forecasting model [129]. Measured high-resolution weather data to capture local microclimate is not easy to achieve at large scales. Data-driven weather forecasting approach became popular to provide more accurate and real-time weather data to UEBM. For example, tools like Urban Weather Generator [130] estimate the hourly urban canopy air temperature and humidity based on rural weather data, e.g., from airports. However, its limitations lie in its coarse resolution and ignoring the interaction between buildings and surrounding environments. Physic-based weather forecasting model remains popular for simulating and predicting the urban microclimate. Particularly, co-simulation between UBEM and UCM is necessary to capture the two-way interactions between buildings and the surrounding urban environment with high resolution and detailed boundary information [7].

In this study, we pioneer an innovative study by evaluating the urban microclimate within the entire San Francisco during a historical heatwave. Our work offers unprecedented granularity in understanding the urban microclimate at a large scale under extreme heat conditions. By leveraging advanced integration between UBEM and UCM tools, our work provides valuable insight into the localized effects of heatwaves on urban air temperature and wind patterns within the city, contributing significantly to our knowledge of urban climate resilience and adaptation strategies.

5.2 Methodology

5.2.1 Microclimate modeling and simulation with CityFFD

CityFFD is a UCM tool that employs semi-Lagrangian and fractional step methods, along with several novel numerical schemes to enhance accuracy and lower computational costs [131], [132], [133]. Its fourth-order numerical interpolation scheme minimizes numerical dissipation and dispersion errors, even when using coarse grids [131]. By using Large Eddy Simulation (LES), CityFFD accurately represents turbulence in the atmospheric boundary layer. Due to its semi-Lagrangian formulation, is unconditionally stable allowing any mesh size or time step. Depending on the domain size and available computational resources, grid resolutions typically range from 1 m to 10 m near buildings, with recommended time steps spanning a few seconds to several minutes for stable and accurate simulations.

5.2.2 Urban building energy modeling with CityBES

CityBES (i.e., City Buildings, Energy, and Sustainability) is one of the UBEM programs [134], [135]. It is a web-based data and computing platform that focuses on energy modeling and analysis of a city's building stock to support district- or city-scale energy efficiency programs [136]. CityBES uses an international open data standard, CityGML, to represent and exchange 3D city models. CityBES employs EnergyPlus to simulate building energy use and savings from energy-efficient retrofits [137]. CityBES provides a suite of features for urban planners, city energy managers, building owners, utilities, energy consultants, and researchers.

5.2.3 Integration between CityFFD and CityBES

The CityBES program utilizes EnergyPlus for simulating building energy usage and assessing energy efficiency retrofits. For each exterior surface of a building, the total heat is calculated as the convective and radiative heat emission rate from the surface, added to the HVAC exhausted and rejected heat emission rate at the zone and building level, subdivided by the surface area of each exterior surface. Hourly total heat emission rates are calculated for each building's exterior surface during the simulation period. Each building exterior surface is mapped to the attaching grid cell(s) in the UCM model. The heat flux into any UCM grid cell is determined by dividing the total heat from attached surfaces by the grid cell's connected area. In addition, this work runs simulations of CityFFD as the UCM tool (Figure 5-5). The CityFFD microclimate model was created for the domain (13,000 meters by 11,800 meters, and 900 meters in height) with a grid size of 10 meters in three dimensions and runs at a 100-second time step. The total grid number of the CityFFD domain is about 138 million. CityFFD simulation was conducted on an Nvidia DGX Station with four Nvidia Tesla V100 Tensor Core GPUs, achieving a total of 500 TFLOPS of performance and offering 128GB of total GPU memory. It took 4 hours for each coupling time step in CityFFD. The ground temperature was set to 40°C based on recent research that derived land surface temperature images from Landsat 8 satellite thermal infrared sensor from 2017 to 2020 in the San Francisco Bay area [138]. The simulation was conducted for the hottest day, September 1st, 2017, and the data was exchanged between CityBES and CityFFD at the hourly time step, which is considered one timestep of the co-simulation.

Figure 5-1 illustrates the general data exchange fields and processes of a couple of UBEM and UCM models. At each iteration, the UBEM provides heat emission and building surface temperature per exterior surface as the boundary conditions for the UCM, while UCM provides local ambient conditions for the UBEM at the CFD cell level, which may include ambient air temperature, humidity, pressure, velocity and wind direction, solar radiation, and carbon dioxide (CO2) level.

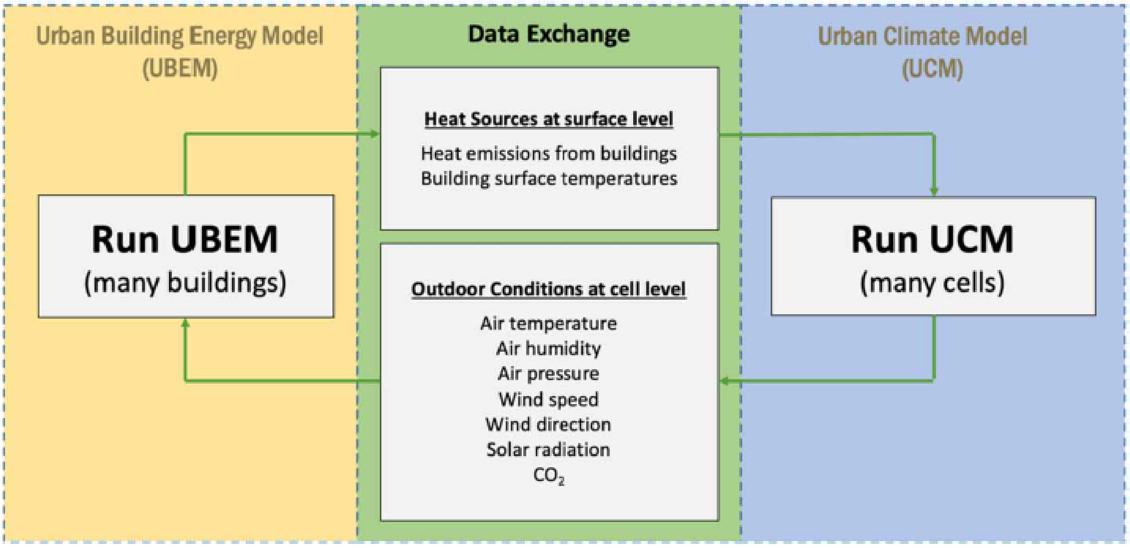


Figure 5-1. Data exchange between UBEM and UCM

A new flexible and tool-agnostic data schema (in JSON) was utilized to facilitate the exchange of data between urban building energy models and urban microclimate models. As shown in Figure 5-2, three JSON files serve as a “bridge” for data exchange between the UCM and UBEM tools. Each UCM grid cell ID is mapped to one or more surface IDs in UBEM. During each coupling iteration, the UCM outputs variables like air temperature, wind speed, and direction (file No. 3), which are mapped to specific buildings and surfaces using file No. 5. Similarly, UBEM outputs, such as temperature and heat emissions, are transferred back to corresponding UCM grid cells via the same mapping. While files No. 3 and No. 4 are dynamically updated during co-simulation, the static JSON files (No. 1, No. 2, and No. 5) are derived from the geographic information of the simulation domain.

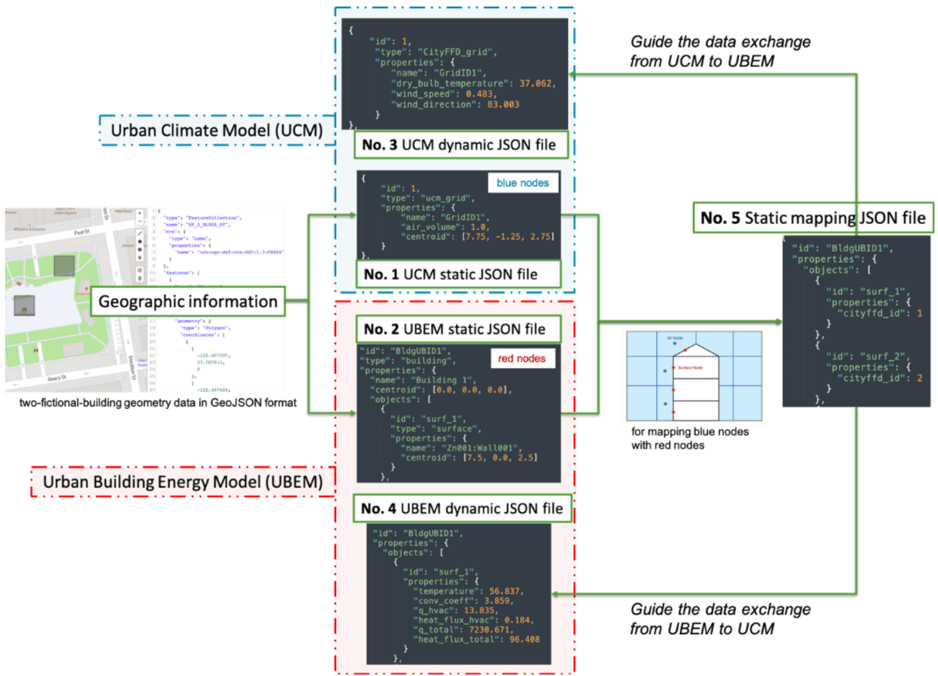


Figure 5-2. Illustration of the data exchange schema using JSON files [86].

Figure 5-3 further illustrates how building surface and ambient air condition data are mapped between air nodes (UCM/CityFFD domain) and surface nodes (UBEM/CityBES domain). The surface nodes (red nodes) represent the exterior surfaces of a UBEM building, such as exterior walls, roofs, and windows. The Air Nodes (blue nodes) are the UCM cells that hold information about the surrounding air conditions. The nearest red nodes and blue nodes were mapped based on the centroid information. During the later coupled simulations, a red node reads outdoor conditions from its mapped blue node, and a blue node takes the aggregated heat flux and surface temperatures from its attached/mapped surfaces as boundary conditions.

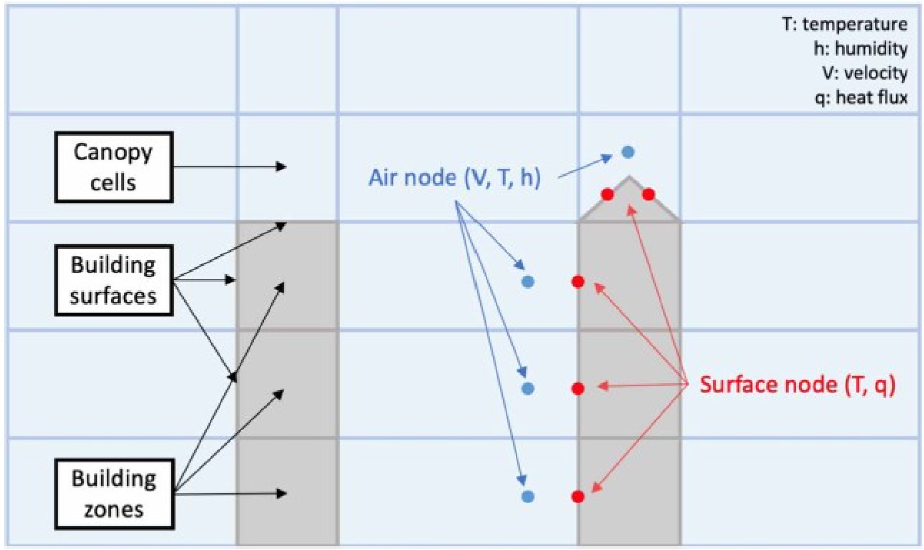


Figure 5-3. Mechanism of mapping the grid cells in UCM (blue nodes) with the building surfaces in UBEM (red nodes) [86].

5.2.4 Case study

In recent years, San Francisco has experienced unprecedented heatwaves during the summer season, impacting both energy demand and public health. Even in cool, coastal California, extreme heat sickens and kills people. In 2017, extreme heat killed 14 people in the San Francisco Bay Area. Over the Labor Day weekend, six people alone died in San Francisco. The heat also sent hundreds more to the hospital [139]. The entire city of San Francisco is selected for the case study to evaluate the urban microclimate during the heatwave by integrated simulation of UBEM and UCM programs. All building information required for the simulation was collected from the public data portal of San Francisco, DataSF [datasf.org]. The dataset comprises 148,689 buildings. The weather data used for UBEM was collected from the local weather stations scattered in the whole domain. San Francisco is located in Northern California, USA, covering an area of 121.4 km^2 (Figure 5-4).

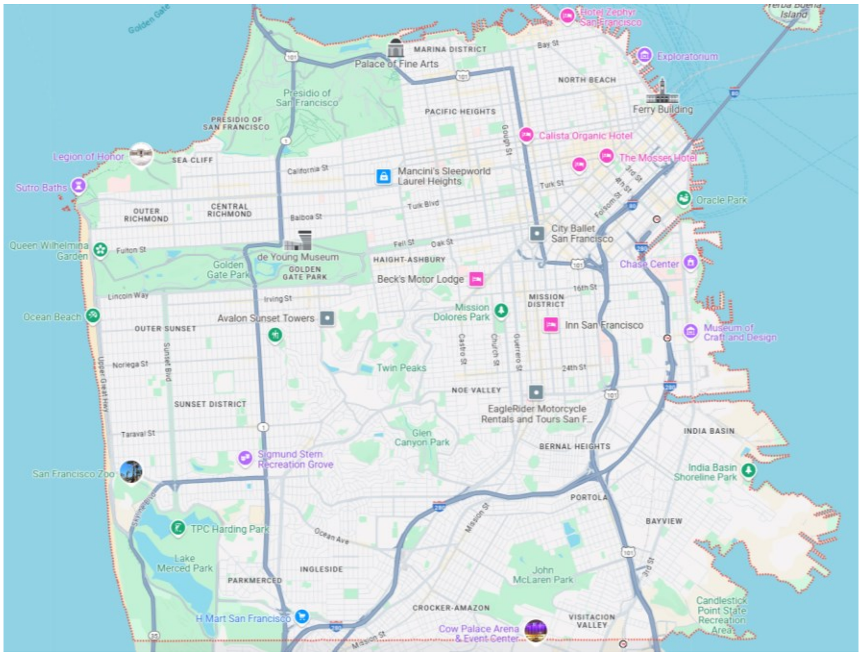


Figure 5-4. Urban layout of San Francisco via Google Maps.

Figure 5-5. CityFFD model of the entire San Francisco city

5.3 Results

Using the data collected from the local weather station installed in San Francisco as a boundary condition, it was observed that at 2 pm on September 1st, 2017, the air temperature was 37.4 °C, and the wind speed was 0.5 m/s from the west. These meteorological data serve as crucial boundary conditions for understanding the local urban microclimate. The heat emissions from buildings during the daytime largely affect the ambient air temperature near buildings. Luo et al. [86] indicate the buildings began to produce heat emissions to the surrounding air starting at 10 am in the San Francisco area, which accelerated the heating of the surrounding air afterward.

5.3.1 Urban temperature distribution

Figure 5-6 illustrates the distribution of ambient air temperature at a height of 10m. The air temperature near buildings increased to 55℃ in most places. In low-density areas, the air temperature remained around 39℃， while high-density areas experienced temperatures exceeding 42 ℃. A notable disparity such as this illustrates the localized impacts of building heat emissions on a city's microclimate. As a result of the high ambient air temperatures in San Francisco, residents are affected by increased building energy consumption and outdoor discomfort (also for tourists).

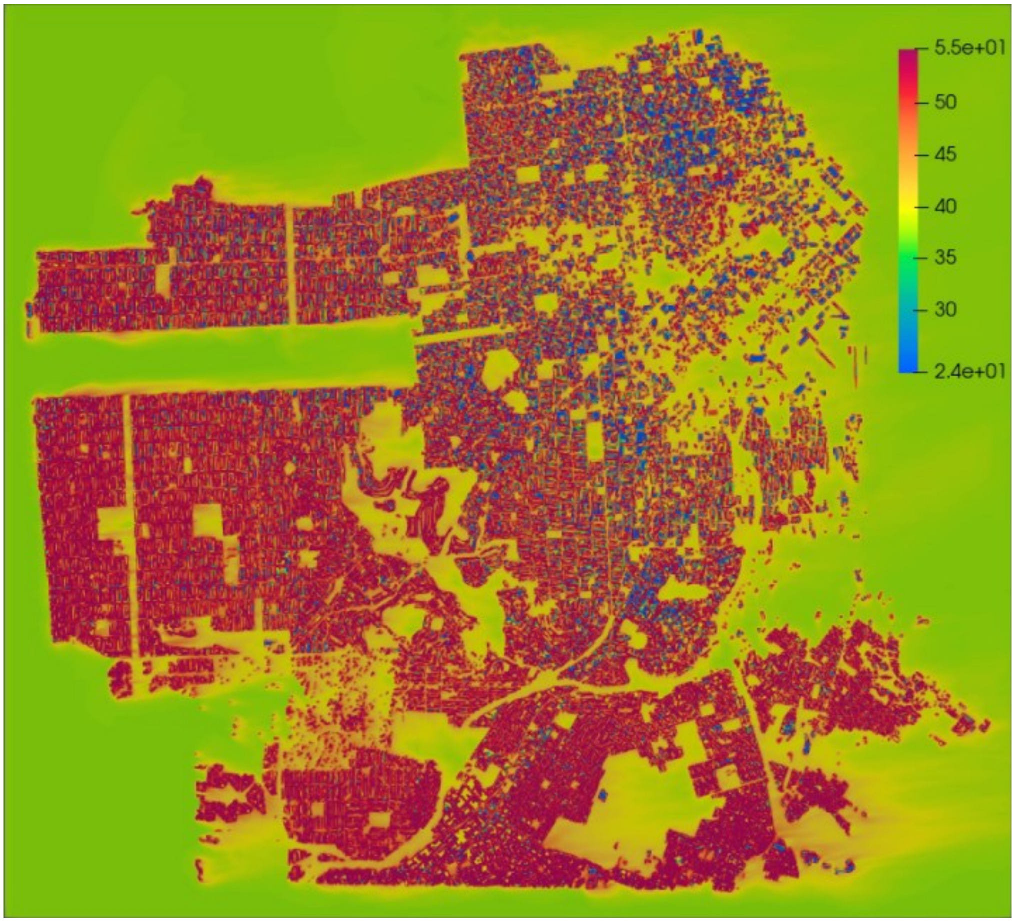


Figure 5-6. San Francisco's air temperature distribution at a height of 10 meters.

5.3.2 Urban wind distribution

Figure 5-7 shows the wind distribution within San Francisco at a height of 10m, specifically at 2 pm on September 1, 2017. As aforementioned, the prevailing wind direction is west, with a speed of 0.5 m/s. However, wind speed reduces to 0.15 m/s in open areas and decreases to zero in densely populated areas. The decreased wind velocity hampers high-temperature air dispersal from urban areas to suburbs. The marked decrease in wind velocity hinders the dispersion of high-temperature air from urban areas into suburban areas, intensifying heatwaves and amplifying the effects of Urban Heat Island (UHI). Due to this, the urban heat island effect gets worse, posing more challenges to mitigating heat-related risks and making sure communities are resilient to heat.

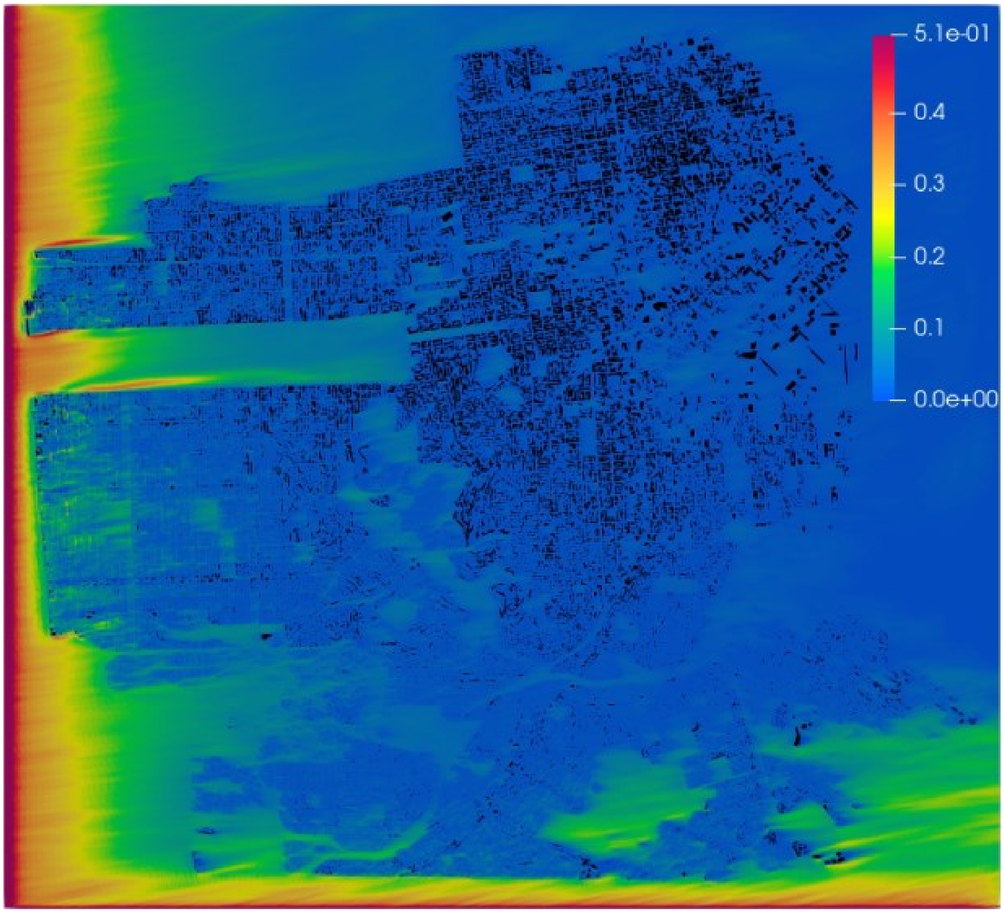


Figure 5-7. San Francisco's wind distribution at a height of 10 meters.

5.4 Summary

Our study utilized an integrated platform to exchange information between UBEM and UCM models with high spatial granularity at a large scale. Using CityFFD and CityBES, we simulated the city's urban microclimate during the hottest hour of September 1st, 2017, a day of record-breaking temperatures. Our findings highlight the significant impact of urban microclimate on ambient air temperature and local wind speed within the city. We observed a considerable increase in ambient air temperature, with the maximum temperature reaching 55°C in dense areas, representing an increased deviation of 8°C from the inlet air temperature at a height of 10m, with the westerly wind originating from the Pacific Ocean. In addition, our findings revealed a significant decrease in wind velocity, which hampers high-temperature air dispersal from urban areas to suburbs (San Francisco Bay). As pioneers in the application of high-granularity urban microclimate modeling at the city scale, our study provides a comprehensive understanding of ambient air temperature and wind velocity in San Francisco during a record heatwave. Our work shows how integrated co-simulation methods can be applied to achieve high-resolution urban microclimate modeling considering detailed dynamic anthropogenic heat from buildings, paving the way for future research and practical applications in urban sustainability and resilience.

Chapter 6. Conclusions

This research has systematically examined the spatiotemporal characterization of urban microclimates and their influence on building-level energy performance by establishing an integrated platform that couples urban microclimate models with building energy simulations. Focusing on a hot and arid climate zone, an area that remains underexplored in existing literature. By replacing conventional rural weather inputs with ambient weather conditions, the coupled modeling framework provides more detailed cooling profiles, helping avoid the common overestimation problems seen in current district cooling system designs and supporting more efficient and context-sensitive planning. In addition, the study explores the application of machine learning (ML) as a surrogate modeling technique for predicting hourly building energy consumption during the design phase. To address the high computational cost associated with generating detailed ML models, a weather clustering-based approach was introduced, effectively reducing the dataset size while maintaining prediction accuracy within ASHRAE's acceptable tolerance. Finally, a city-scale integration method was formulated to assess urban microclimate conditions, particularly air temperature and wind distribution, during extreme heat events. Overall, these contributions enhance the assessment of urban microclimates and support more efficient and accurate building energy planning at various scales.

6.1 Contributions

The following is a list of significant findings and contributions from this thesis:

- Developed a novel coupling strategy that integrates a CFD-based urban microclimate model with a building energy model, enabling detailed spatiotemporal evaluation of urban microclimate effects on building-level energy performance.
- Conducted a comprehensive global sensitivity analysis to explore the complex relationship

between input parameters and building energy consumption, identifying outdoor air temperature as the most influential factor, followed by wind speed and cooling setpoint.

- Proposed an innovative machine learning-based approach for hourly building energy prediction by applying a weather clustering technique. This method effectively reduces computational demand by selecting a representative subset of days from annual weather data.
- Established a coupling platform linking urban microclimate models and urban building energy models using a JSON-based data exchange schema. The platform supports scalable analysis of building-environment interactions at the city scale.

6.2 Future work

In the context of future work for this thesis, a range of opportunities exist to advance and refine the current research:

- Continue exploring urban microclimate impacts on large-scale energy performance by applying the proposed UCM-UBEM platform.
- Incorporate relative humidity into urban microclimate evaluations, to better understand how varying moisture levels affect building cooling and heating demands.
- Apply the developed ML-based approaches at different spatial scales, from neighborhoods or districts to entire urban areas, for more detailed insights into urban energy consumption patterns.
- Adopt a multi-zone configuration to better capture intra-floor thermal variations, occupancy diversity, and localized HVAC control for more accurate energy simulations.

Addressing these areas in future research will further enhance our understanding of urban microclimates and enable more accurate assessments of building energy performance across different scales.

REFERENCES

[1] U. Ranhagen and K. Groth, *The symbiocity approach: a conceptual framework for sustainable urban development*. SKL International, 2012.

[2] E. Prataviera, J. Vivian, G. Lombardo, and A. Zarrella, 'Evaluation of the impact of input uncertainty on urban building energy simulations using uncertainty and sensitivity analysis', *Applied Energy*, vol. 311, p. 118691, Apr. 2022, doi: 10.1016/j.apenergy.2022.118691.

[3] F. Kuznik, L. Frayssinet, J.-J. Roux, and L. Merlier, 'Calculation of heating and cooling energy loads at the district scale: Development of MoDEM, a modular and technologically explicit platform', *Sustainable Cities and Society*, vol. 83, p. 103901, Aug. 2022, doi: 10.1016/j.scs.2022.103901.

[4] D. Coakley, P. Raftery, and M. Keane, 'A review of methods to match building energy simulation models to measured data', *Renewable and Sustainable Energy Reviews*, vol. 37, pp. 123–141, Sep. 2014, doi: 10.1016/j.rser.2014.05.007.

[5] Y. Q. Ang, Z. M. Berzolla, and C. F. Reinhart, 'From concept to application: A review of use cases in urban building energy modeling', *Applied Energy*, vol. 279, p. 115738, Dec. 2020, doi: 10.1016/j.apenergy.2020.115738.

[6] M. Hadavi and H. Pasdarshahri, 'Impacts of urban buildings on microclimate and cooling systems efficiency: Coupled CFD and BES simulations', *Sustainable Cities and Society*, vol. 67, p. 102740, Jan. 2021, doi: 10.1016/j.scs.2021.102740.

[7] X. S. Yang, L. H. Zhao, M. Bruse, and Q. L. Meng, 'Assessing the Effect of Microclimate on Building Energy Performance by Co-Simulation', *AMM*, vol. 121–126, pp. 2860–2867, Oct. 2011, doi: 10.4028/www.scientific.net/AMM.121-126.2860.

[8] R. Abrar *et al.*, 'Assessing the Spatial Mapping of Heat Vulnerability under Urban Heat Island (UHI) Effect in the Dhaka Metropolitan Area', *Sustainability*, vol. 14, no. 9, Art. no. 9, Jan. 2022, doi: 10.3390/su14094945.

[9] A. Sedaghat and M. Sharif, 'Mitigation of the impacts of heat islands on energy consumption in buildings: A case study of the city of Tehran, Iran', *Sustainable Cities and Society*, vol. 76, p. 103435, Jan. 2022, doi: 10.1016/j.scs.2021.103435.

[10] J. Litardo *et al.*, 'Urban Heat Island intensity and buildings' energy needs in Duran, Ecuador: Simulation studies and proposal of mitigation strategies', *Sustainable Cities and Society*, vol. 62, p. 102387, Nov. 2020, doi: 10.1016/j.scs.2020.102387.

[11] M. Singh and R. Sharston, 'Quantifying the dualistic nature of urban heat Island effect (UHI) on building energy consumption', *Energy and Buildings*, vol. 255, p. 111649, Jan. 2022, doi: 10.1016/j.enbuild.2021.111649.

[12] X. Li, Y. Zhou, S. Yu, G. Jia, H. Li, and W. Li, 'Urban heat island impacts on building energy consumption: A review of approaches and findings', *Energy*, vol. 174, pp. 407–419, May 2019, doi: 10.1016/j.energy.2019.02.183.

[13] M. Zinzi, E. Carnielo, and B. Mattoni, 'On the relation between urban climate and energy performance of buildings. A three-years experience in Rome, Italy', *Applied Energy*, vol. 221, pp. 148–160, Jul. 2018, doi: 10.1016/j.apenergy.2018.03.192.

[14] M. Santamouris, 'On the energy impact of urban heat island and global warming on buildings', *Energy and Buildings*, vol. 82, pp. 100–113, Oct. 2014, doi: 10.1016/j.enbuild.2014.07.022.

[15] M. Palme, L. Inostroza, G. Villacreses, A. Lobato-Cordero, and C. Carrasco, 'From urban climate to energy consumption. Enhancing building performance simulation by including the urban heat island effect', *Energy and Buildings*, vol. 145, pp. 107–120, Jun. 2017, doi: 10.1016/j.enbuild.2017.03.069.

[16] C. Li *et al.*, 'Interaction between urban microclimate and electric air-conditioning energy consumption during high temperature season', *Applied Energy*, vol. 117, pp. 149–156, Mar. 2014, doi: 10.1016/j.apenergy.2013.11.057.

[17] Y. Toparlar, B. Blocken, B. Maiheu, and G. J. F. van Heijst, 'Impact of urban microclimate on summertime building cooling demand: A parametric analysis for Antwerp, Belgium', *Applied Energy*, vol. 228, pp. 852–872, Oct. 2018, doi: 10.1016/j.apenergy.2018.06.110.

[18] K. Javanroodi and V. M. Nik, 'Impacts of microclimate conditions on the energy performance of buildings in urban areas', *Buildings*, vol. 9, no. 8, Aug. 2019, doi: 10.3390/buildings9080189.

[19] H. Akbari and D. Kolokotsa, 'Three decades of urban heat islands and mitigation technologies research', *Energy and Buildings*, vol. 133, pp. 834–842, Dec. 2016, doi: 10.1016/j.enbuild.2016.09.067.

[20] M. Santamouris, C. Cartalis, A. Synnefa, and D. Kolokotsa, 'On the impact of urban heat island and global warming on the power demand and electricity consumption of buildings—A review', *Energy and Buildings*, vol. 98, pp. 119–124, Jul. 2015, doi: 10.1016/j.enbuild.2014.09.052.

[21] B. Ameer and M. Krarti, 'Review of Urban Heat Island and Building Energy Modeling Approaches', *ASME Journal of Engineering for Sustainable Buildings and Cities*, vol. 3, no. 1,

Feb. 2022, doi: 10.1115/1.4053677.

[22] J. Kong, Y. Zhao, J. Carmeliet, and C. Lei, 'Urban Heat Island and Its Interaction with Heatwaves: A Review of Studies on Mesoscale', *Sustainability*, vol. 13, no. 19, Art. no. 19, Jan. 2021, doi: 10.3390/su131910923.

[23] J. Unger, 'Intra-urban relationship between surface geometry and urban heat island: review and new approach', *CLIMATE RESEARCH*, vol. 27, pp. 253–264, 2004.

[24] R. Aghamolaei, M. Fallahpour, and P. A. Mirzaei, 'Tempo-spatial thermal comfort analysis of urban heat island with coupling of CFD and building energy simulation', *Energy and Buildings*, vol. 251, p. 111317, Nov. 2021, doi: 10.1016/j.enbuild.2021.111317.

[25] E. J. Gago, J. Roldan, R. Pacheco-Torres, and J. Ordóñez, 'The city and urban heat islands: A review of strategies to mitigate adverse effects', *Renewable and Sustainable Energy Reviews*, vol. 25, pp. 749–758, Sep. 2013, doi: 10.1016/j.rser.2013.05.057.

[26] K. Degirmenci, K. C. Desouza, W. Fieuw, R. T. Watson, and T. Yigitcanlar, 'Understanding policy and technology responses in mitigating urban heat islands: A literature review and directions for future research', *Sustainable Cities and Society*, vol. 70, p. 102873, Jul. 2021, doi: 10.1016/j.scs.2021.102873.

[27] N. Sezer, H. Yoonus, D. Zhan, L. (Leon) Wang, I. G. Hassan, and M. A. Rahman, 'Urban microclimate and building energy models: A review of the latest progress in coupling strategies', *Renewable and Sustainable Energy Reviews*, vol. 184, p. 113577, Sep. 2023, doi: 10.1016/j.rser.2023.113577.

[28] J. Liu, M. Heidarinejad, S. K. Nikkho, N. W. Mattise, and J. Srebric, 'Quantifying impacts of urban microclimate on a building energy consumption—a case study', *Sustainability*, vol. 11, no. 18, p. 4921, 2019.

[29] X. Yang, L. Zhao, M. Bruse, and Q. Meng, 'An integrated simulation method for building energy performance assessment in urban environments', *Energy and Buildings*, vol. 54, pp. 243–251, Nov. 2012, doi: 10.1016/j.enbuild.2012.07.042.

[30] J. Allegrini and J. Carmeliet, 'Simulations of local heat islands in Zürich with coupled CFD and building energy models', *Urban Climate*, vol. 24, pp. 340–359, Jun. 2018, doi: 10.1016/j.uclim.2017.02.003.

[31] K. Gobakis and D. Kolokotsa, 'Coupling building energy simulation software with microclimatic simulation for the evaluation of the impact of urban outdoor conditions on the energy

consumption and indoor environmental quality’, *Energy and Buildings*, vol. 157, pp. 101–115, Dec. 2017, doi: 10.1016/j.enbuild.2017.02.020.

[32] J. Bouyer, C. Inard, and M. Musy, ‘Microclimatic coupling as a solution to improve building energy simulation in an urban context’, *Energy and Buildings*, vol. 43, no. 7, pp. 1549–1559, Jul. 2011, doi: 10.1016/j.enbuild.2011.02.010.

[33] L. Malys, M. Musy, and C. Inard, ‘Microclimate and building energy consumption: study of different coupling methods’, *Advances in Building Energy Research*, vol. 9, no. 2, pp. 151–174, Jul. 2015, doi: 10.1080/17512549.2015.1043643.

[34] Y. K. Yi and N. Feng, ‘Dynamic integration between building energy simulation (BES) and computational fluid dynamics (CFD) simulation for building exterior surface’, *Build. Simul.*, vol. 6, no. 3, pp. 297–308, Sep. 2013, doi: 10.1007/s12273-013-0116-9.

[35] Y. K. Yi and A. M. Malkawi, ‘Integrating neural network models with computational fluid dynamics (CFD) for site-specific wind condition’, *Building Simulation*, vol. 4, no. 3, pp. 245–254, Sep. 2011, doi: 10.1007/s12273-011-0042-7.

[36] K. Amasyali and N. El-Gohary, ‘Machine learning for occupant-behavior-sensitive cooling energy consumption prediction in office buildings’, *Renewable and Sustainable Energy Reviews*, vol. 142, p. 110714, May 2021, doi: 10.1016/j.rser.2021.110714.

[37] K. Li, C. Hu, G. Liu, and W. Xue, ‘Building’s electricity consumption prediction using optimized artificial neural networks and principal component analysis’, *Energy and Buildings*, vol. 108, pp. 106–113, Dec. 2015, doi: 10.1016/j.enbuild.2015.09.002.

[38] R. Platon, V. R. Dehkordi, and J. Martel, ‘Hourly prediction of a building’s electricity consumption using case-based reasoning, artificial neural networks and principal component analysis’, *Energy and Buildings*, vol. 92, pp. 10–18, Apr. 2015, doi: 10.1016/j.enbuild.2015.01.047.

[39] Z. Wang, T. Hong, and M. A. Piette, ‘Building thermal load prediction through shallow machine learning and deep learning’, *Applied Energy*, vol. 263, p. 114683, Apr. 2020, doi: 10.1016/j.apenergy.2020.114683.

[40] Y. Yue, Z. Yan, P. Ni, F. Lei, and G. Qin, ‘Promoting solar energy utilization: Prediction, analysis and evaluation of solar radiation on building surfaces at city scale’, *Energy and Buildings*, vol. 319, p. 114561, Sep. 2024, doi: 10.1016/j.enbuild.2024.114561.

[41] J. Wu *et al.*, ‘A Comparative Analysis of Machine Learning-Based Energy Baseline Models

across Multiple Building Types', *Energies*, vol. 17, no. 6, 2024, doi: 10.3390/en17061285.

[42] D. Ma, X. Li, B. Lin, Y. Zhu, and S. Yue, 'A dynamic intelligent building retrofit decision-making model in response to climate change', *Energy and Buildings*, vol. 284, 2023, doi: 10.1016/j.enbuild.2023.112832.

[43] T. Ben Hassen, H. El Bilali, and M. Al-Maadeed, 'Agri-Food Markets in Qatar: Drivers, Trends, and Policy Responses', *Sustainability*, vol. 12, no. 9, p. 3643, May 2020, doi: 10.3390/su12093643.

[44] E. Zavrl, M. E. Mankibi, M. Dovjak, and U. Stritih, 'Experimental Investigation of Air-based Active-Passive System for Cooling Application in Buildings', *Sustainable Cities and Society*, p. 104031, Jun. 2022, doi: 10.1016/j.scs.2022.104031.

[45] P. M. Rathnam and S. Ram, 'Imbibing Energy Efficiency in Buildings Through Sustainable Materials—A Review', in *Advances in Sustainable Materials and Resilient Infrastructure*, K. R. Reddy, R. K. Pancharathi, N. G. Reddy, and S. R. Arukala, Eds., in Springer Transactions in Civil and Environmental Engineering. , Singapore: Springer, 2022, pp. 399–414. doi: 10.1007/978-981-16-9744-9_27.

[46] M. Indraganti, D. Boussaa, S. Assadi, and E. Mostavi, 'User satisfaction and energy use behavior in offices in Qatar', *Building Services Engineering Research and Technology*, vol. 39, no. 4, pp. 391–405, Jul. 2018, doi: 10.1177/0143624417751388.

[47] S. W. Kim and R. D. Brown, 'Urban heat island (UHI) intensity and magnitude estimations: A systematic literature review', *Science of The Total Environment*, vol. 779, p. 146389, Jul. 2021, doi: 10.1016/j.scitotenv.2021.146389.

[48] M. W. Rotach *et al.*, 'BUBBLE – an Urban Boundary Layer Meteorology Project', *Theor. Appl. Climatol.*, vol. 81, no. 3, pp. 231–261, Jul. 2005, doi: 10.1007/s00704-004-0117-9.

[49] M. I. A. S. Al-Marzooqi, 'The Impact of Urban Heat Island (UHI) Phenomenon on the Climate of Qatar', M.S., Hamad Bin Khalifa University (Qatar), Qatar.

[50] L. Shi, Z. Luo, W. Matthews, Z. Wang, Y. Li, and J. Liu, 'Impacts of urban microclimate on summertime sensible and latent energy demand for cooling in residential buildings of Hong Kong', *Energy*, vol. 189, p. 116208, Dec. 2019, doi: 10.1016/j.energy.2019.116208.

[51] M. Santamouris and D. Kolokotsa, *Urban Climate Mitigation Techniques*. Routledge, 2016.

[52] B.-J. He, W. Wang, A. Sharifi, and X. Liu, 'Progress, knowledge gap and future directions of urban heat mitigation and adaptation research through a bibliometric review of history and

evolution', *Energy and Buildings*, vol. 287, p. 112976, May 2023, doi: 10.1016/j.enbuild.2023.112976.

[53] K. Marthya, R. Furlan, L. Ellath, M. Esmat, and R. Al-Matwi, 'Place-Making of Transit Towns in Qatar: The Case of Qatar National Museum-Souq Waqif Corridor', *Designs*, vol. 5, no. 1, p. 18, Mar. 2021, doi: 10.3390/designs5010018.

[54] L. A. Ellath, 'URBAN FORM AND MOBILITY ANALYSIS IN SUSTAINABLE CITIES: CASE OF MARINA DISTRICT, LUSAIL CITY', Qatar University, Qatar, 2022.

[55] C. Méndez and Y. Bicer, 'Qatar's Wind Energy Potential with Associated Financial and Environmental Benefits for the Natural Gas Industry', *Energies*, vol. 12, no. 17, Art. no. 17, Jan. 2019, doi: 10.3390/en12173329.

[56] B. Jia, D. Hou, A. Kamal, I. G. Hassan, and L. Wang, 'Developing machine-learning meta-models for high-rise residential district cooling in hot and humid climate', *Journal of Building Performance Simulation*, vol. 15, no. 4, pp. 553–573, Jul. 2022, doi: 10.1080/19401493.2021.2001573.

[57] M. M. Mohamed and M. H. Almarshadi, 'Simulation of District Cooling Plant and Efficient Energy Air Cooled Condensers (Part I)', *Journal of Electronics Cooling and Thermal Control*, vol. 07, no. 03, Art. no. 03, Jul. 2017, doi: 10.4236/jectc.2017.73005.

[58] P. Shen, M. Dai, P. Xu, and W. Dong, 'Building heating and cooling load under different neighbourhood forms: Assessing the effect of external convective heat transfer', *Energy*, vol. 173, pp. 75–91, Apr. 2019, doi: 10.1016/j.energy.2019.02.062.

[59] B. Czernecki, A. Głogowski, and J. Nowosad, 'Climate: An R Package to Access Free In-Situ Meteorological and Hydrological Datasets For Environmental Assessment', *Sustainability*, vol. 12, no. 1, Art. no. 1, Jan. 2020, doi: 10.3390/su12010394.

[60] A. Katal, M. Mortezazadeh, and L. (Leon) Wang, 'Modeling building resilience against extreme weather by integrated CityFFD and CityBEM simulations', *Applied Energy*, vol. 250, pp. 1402–1417, Sep. 2019, doi: 10.1016/j.apenergy.2019.04.192.

[61] M. Mortezazadeh Dorostkar, 'CityFFD – City Fast Fluid Dynamics Model for Urban Microclimate Simulations', phd, Concordia University, 2019. Accessed: Jul. 06, 2022. [Online]. Available: https://spectrum.library.concordia.ca/id/eprint/985976/

[62] 'Solving city and building microclimates by fast fluid dynamics with large timesteps and coarse meshes - ScienceDirect'. Accessed: Mar. 02, 2023. [Online]. Available:

https://www.sciencedirect.com/science/article/pii/S0360132320303140

[63] A. Katal, M. Mortezazadeh, and L. L. Wang, ‘Modeling building resilience against extreme weather by integrated CityFFD and CityBEM simulations’, *Applied Energy*, vol. 250, pp. 1402–1417, 2019, doi: 10.1016/j.apenergy.2019.04.192.

[64] U. S. N. A. and S. Administration, *U.S. Standard Atmosphere, 1976*. National Oceanic and Atmospheric Administration, 1976.

[65] R. and A.-C. E. American Society of Heating, *2005 ASHRAE handbook: fundamentals.* Atlanta, GA.: ASHRAE, 2005.

[66] ‘Outside Surface Heat Balance: Engineering Reference — EnergyPlus 8.7’. Accessed: May 12, 2022. [Online]. Available: https://bigladdersoftware.com/epx/docs/8-7/engineering-reference/outside-surface-heat-balance.html

[67] P. G. Ellis and P. A. Torcellini, ‘Simulating Tall Buildings Using EnergyPlus: Preprint’, p. 12.

[68] W. Tian, ‘A review of sensitivity analysis methods in building energy analysis’, *Renewable and Sustainable Energy Reviews*, vol. 20, pp. 411–419, Apr. 2013, doi: 10.1016/j.rser.2012.12.014.

[69] ‘Lusail City GSAS 2 Star Rating Guidelines.pdf’. Accessed: Feb. 11, 2022. [Online]. Available: https://www.lusail.com/wp-content/uploads/LCAC%20Files/Section%2003%20Building%20Permit%20Application%20Procedures/GSAS%20Toolkit%20and%20Calculator%20Pack%20for%202%20Star%20Rating/Lusail%20City%20GSAS%202%20Star%20Rating%20Guidelines.pdf

[70] G. Manache and C. S. Melching, ‘Sensitivity Analysis of a Water-Quality Model Using Latin Hypercube Sampling’, *J. Water Resour. Plann. Manage.*, vol. 130, no. 3, pp. 232–242, May 2004, doi: 10.1061/(ASCE)0733-9496(2004)130:3(232).

[71] H. Jia and A. Chong, ‘eplusr: A framework for integrating building energy simulation and data-driven analytics’, *Energy and Buildings*, vol. 237, p. 110757, Apr. 2021, doi: 10.1016/j.enbuild.2021.110757.

[72] ‘ANSI/ASHRAE/IES Standard 90.1-2019 -- Energy Standard for Building Except Low-Rise Residential Buildings’. Accessed: Sep. 15, 2022. [Online]. Available: https://www.ashrae.org/technical-resources/bookstore/standard-90-1

[73] ‘EN 13363-2:2005/AC:2006 - Solar protection devices combined with glazing - Calculation of total solar energy transmittance and light transmittance - Part 2: Detailed calculation method’, iTeh Standards. Accessed: Mar. 02, 2023. [Online]. Available:

https://standards.iteh.ai/catalog/standards/cen/88ae67b1-7af0-40d7-a368-59e44bbc5194/en-13363-2-2005-ac-2006

[74] 'Qatar Construction Specification'. Accessed: Dec. 10, 2022. [Online]. Available: http://shared4.info/architecture/qcs/index.php?option=com_content&view=category&id=22%20

[75] H. Lim and Z. J. Zhai, 'Comprehensive evaluation of the influence of meta-models on Bayesian calibration', *Energy and Buildings*, vol. 155, pp. 66–75, 2017.

[76] D. Hou, I. G. Hassan, and L. Wang, 'Review on building energy model calibration by Bayesian inference', *Renewable and Sustainable Energy Reviews*, vol. 143, p. 110930, 2021.

[77] J. Yang, H. Fu, and M. Qin, 'Evaluation of Different Thermal Models in EnergyPlus for Calculating Moisture Effects on Building Energy Consumption in Different Climate Conditions', *Procedia Engineering*, vol. 121, pp. 1635–1641, 2015, doi: 10.1016/j.proeng.2015.09.194.

[78] A. Salam *et al.*, 'Meteorological Factors and Seasonal Stroke Rates: A Four-year Comprehensive Study', *Journal of Stroke and Cerebrovascular Diseases*, vol. 28, no. 8, pp. 2324–2331, Aug. 2019, doi: 10.1016/j.jstrokecerebrovasdis.2019.05.032.

[79] I. Kovacic and V. Zoller, 'Building life cycle optimization tools for early design phases', *Energy*, vol. 92, pp. 409–419, Dec. 2015, doi: 10.1016/j.energy.2015.03.027.

[80] R. E. Edwards, J. New, and L. E. Parker, 'Predicting future hourly residential electrical consumption: A machine learning case study', *Energy and Buildings*, vol. 49, pp. 591–603, Jun. 2012, doi: 10.1016/j.enbuild.2012.03.010.

[81] F. Martellotta, U. Ayr, A. Cannavale, S. Liuzzi, and C. Rubino, 'Using neural networks to predict hourly energy consumptions in office and industrial buildings as a function of weather data', *J. Phys.: Conf. Ser.*, vol. 2385, no. 1, p. 012097, Dec. 2022, doi: 10.1088/1742-6596/2385/1/012097.

[82] J. Pegues, 'THE BENEFITS OF 8760 HOUR-BY-HOUR BUILDING ENERGY ANALYSIS'.

[83] S. Seyedzadeh, F. P. Rahimian, I. Glesk, and M. Roper, 'Machine learning for estimation of building energy consumption and performance: a review', *Visualization in Engineering*, vol. 6, no. 1, p. 5, Oct. 2018, doi: 10.1186/s40327-018-0064-7.

[84] K. Ding, J. Kaiser Calautit, and C. Jimenez-Bescos, 'Significance of external wind conditions on the convective heat transfer coefficients (CHTC) and energy performance in multi-zone high-rise buildings', *Energy and Buildings*, vol. 320, p. 114570, Oct. 2024, doi: 10.1016/j.enbuild.2024.114570.

[85] D. B. Crawley *et al.*, 'EnergyPlus: creating a new-generation building energy simulation program', *Energy and Buildings*, vol. 33, no. 4, pp. 319–331, Apr. 2001, doi: 10.1016/S0378-7788(00)00114-6.

[86] N. Luo *et al.*, 'A data schema for exchanging information between urban building energy models and urban microclimate models in coupled simulations', *Journal of Building Performance Simulation*, vol. 0, no. 0, pp. 1–18, Nov. 2022, doi: 10.1080/19401493.2022.2142295.

[87] D. Zhan, N. Sezer, I. G. Hassan, and L. {Leon} Wang, 'Comparative analysis of uncertainty characterization methods in urban building energy models in hot-arid regions', presented at the Building Simulation 2023, in Building Simulation, vol. 18. IBPSA, 2023, pp. 3840–3846. doi: 10.26868/25222708.2023.1725.

[88] N. Nazarian, J. A. Acero, and L. Norford, 'Outdoor thermal comfort autonomy: Performance metrics for climate-conscious urban design', *Building and Environment*, vol. 155, pp. 145–160, 2019, doi: 10.1016/j.buildenv.2019.03.028.

[89] L. Álvarez Flórez, T. Péan, and J. Salom, 'Hourly based methods to assess carbon footprint flexibility and primary energy use in decarbonized buildings', *Energy and Buildings*, vol. 294, p. 113213, Sep. 2023, doi: 10.1016/j.enbuild.2023.113213.

[90] J. G. Bartzis *et al.*, 'A novel approach for air quality trend studies and its application to European urban environments', presented at the 19th International Conference on Harmonisation within Atmospheric Dispersion Modelling for Regulatory Purposes, Harmo 2019, 2019. [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85084160267&partnerID=40&md5=1d2c2ff3956bd60ec03218c200d93921

[91] F. Souayfane, R. M. Lima, H. Dahrouj, and O. Knio, 'A weather-clustering and energy-thermal comfort optimization methodology for indoor cooling in subtropical desert climates', *Journal of Building Engineering*, vol. 51, p. 104327, Jul. 2022, doi: 10.1016/j.jobe.2022.104327.

[92] R. Liao, G. Li, S. Miao, Y. Lu, J. Zhu, and L. Shen, 'Weather-clustering based strategy design for dynamic demand response building HVAC control', presented at the BuildSys 2012 - Proceedings of the 4th ACM Workshop on Embedded Systems for Energy Efficiency in Buildings, 2012, pp. 33–35. doi: 10.1145/2422531.2422537.

[93] I. A. Klampanos, A. Davvetas, S. Andronopoulos, C. Pappas, A. Ikonomopoulos, and V. Karkaletsis, 'Autoencoder-driven weather clustering for source estimation during nuclear events', *Environmental Modelling & Software*, vol. 102, pp. 84–93, Apr. 2018, doi:

10.1016/j.envsoft.2018.01.014.

[94] P. Hoffmann and K. H. Schlünzen, ‘Weather Pattern Classification to Represent the Urban Heat Island in Present and Future Climate’, *Journal of Applied Meteorology and Climatology*, vol. 52, no. 12, pp. 2699–2714, Dec. 2013, doi: 10.1175/JAMC-D-12-065.1.

[95] A. C. Onal, O. Berat Sezer, M. Ozbayoglu, and E. Dogdu, ‘Weather data analysis and sensor fault detection using an extended IoT framework with semantics, big data, and machine learning’, presented at the Proceedings - 2017 IEEE International Conference on Big Data, Big Data 2017, 2017, pp. 2037–2046. doi: 10.1109/BigData.2017.8258150.

[96] J. A. Acero, E. J. K. Koh, G. Pignatta, and L. K. Norford, ‘Clustering weather types for urban outdoor thermal comfort evaluation in a tropical area’, *Theor Appl Climatol*, vol. 139, no. 1–2, pp. 659–675, Jan. 2020, doi: 10.1007/s00704-019-02992-9.

[97] M. Yashaswini, G. Sowmya, and S. M. Ilias, ‘WEATHER CLUSTERING USING K-MEANS’, vol. 9, no. 6, 2021.

[98] J. F. Lundell, B. Bean, and J. Symanzik, ‘Let’s talk about the weather: a cluster-based approach to weather forecast accuracy’, *Comput Stat*, vol. 38, no. 3, pp. 1135–1155, Sep. 2023, doi: 10.1007/s00180-023-01339-3.

[99] B. A. Babanov, V. A. Semenov, and I. I. Mokhov, ‘Comparison of Cluster Analysis Methods for Identifying Weather Regimes in the Euro-Atlantic Region for Winter and Summer Seasons’, *Izv. Atmos. Ocean. Phys.*, vol. 59, no. 6, pp. 605–623, Dec. 2023, doi: 10.1134/S0001433823060026.

[100] X. J. Luo, ‘A novel clustering-enhanced adaptive artificial neural network model for predicting day-ahead building cooling demand’, *Journal of Building Engineering*, vol. 32, p. 101504, Nov. 2020, doi: 10.1016/j.jobe.2020.101504.

[101] C. Peng, Z. Li, Q. Xu, X. Li, X. Li, and H. Chen, ‘Spatial distribution of energy consumption: Integrating climate and macro-statistics for insights from clustering and sensitivity analysis’, *Energy and Buildings*, vol. 318, p. 114446, Sep. 2024, doi: 10.1016/j.enbuild.2024.114446.

[102] L. Gao, T. Liu, T. Cao, Y. Hwang, and R. Radermacher, ‘Comparing deep learning models for multi energy vectors prediction on multiple types of building’, *Applied Energy*, vol. 301, p. 117486, Nov. 2021, doi: 10.1016/j.apenergy.2021.117486.

[103] ‘ASHRAE Standard 62.1 - 2019’. Accessed: Jun. 08, 2022. [Online]. Available: https://ashrae.iwrapper.com/ASHRAE_PREVIEW_ONLY_STANDARDS/STD_62.1_2019

[104] Y. Alhorr and E. Elsarrag, ‘Climate Change Mitigation through Energy Benchmarking in the

GCC Green Buildings Codes', *Buildings*, vol. 5, no. 2, Art. no. 2, Jun. 2015, doi: 10.3390/buildings5020700.

[105] C. Cerezo, J. Sokol, S. AlKhaled, C. Reinhart, A. Al-Mumin, and A. Hajiah, 'Comparison of four building archetype characterization methods in urban building energy modeling (UBEM): A residential case study in Kuwait City', *Energy and Buildings*, vol. 154, pp. 321–334, Nov. 2017, doi: 10.1016/j.enbuild.2017.08.029.

[106] M. Krarti, F. Ali, A. Alaidroos, and M. Houchati, 'Macro-economic benefit analysis of large scale building energy efficiency programs in Qatar', *International Journal of Sustainable Built Environment*, vol. 6, no. 2, pp. 597–609, Dec. 2017, doi: 10.1016/j.ijsbe.2017.12.006.

[107] S. Ray and R. H. Turi, 'Determination of Number of Clusters in K-Means Clustering and Application in Colour Image Segmentation'.

[108] 'k-Medoids - RapidMiner Documentation'. Accessed: Jul. 31, 2023. [Online]. Available: https://docs.rapidminer.com/latest/studio/operators/modeling/segmentation/k_medoids.html

[109] Z. Liu and R. George, 'Mining Weather Data Using Fuzzy Cluster Analysis', in *Fuzzy Modeling with Spatial Information for Geographic Problems*, F. E. Petry, V. B. Robinson, and M. A. Cobb, Eds., Berlin, Heidelberg: Springer, 2005, pp. 105–119. doi: 10.1007/3-540-26886-3_5.

[110] J. Shen, S. I. Chang, E. S. Lee, Y. Deng, and S. J. Brown, 'Determination of cluster number in clustering microarray data', *Applied Mathematics and Computation*, vol. 169, no. 2, pp. 1172–1185, Oct. 2005, doi: 10.1016/j.amc.2004.10.076.

[111] D. Hou and R. Evins, 'A protocol for developing and evaluating neural network-based surrogate models and its application to building energy prediction', *Renewable and Sustainable Energy Reviews*, vol. 193, p. 114283, Apr. 2024, doi: 10.1016/j.rser.2024.114283.

[112] D. Hou *et al.*, 'Detailed Analysis of Microclimate and its Impact on Thermal and Energy Performance of High-Rise Buildings: A Case Study', in *ASHRAE Topical Conference Proceedings*, Atlanta, United States: American Society of Heating, Refrigeration and Air Conditioning Engineers, Inc., 2022, pp. 1–8. Accessed: Nov. 10, 2023. [Online]. Available: https://www.proquest.com/docview/2696514312/abstract/705A26F26F7A4642PQ/1

[113] Dongxue Zhan, Danlin Hou, Liangzhu (Leon) Wang, and I. G. Hassan, 'Comparing Two Simulation Approaches for Quantifying Urban Microclimate Impact on Building Energy Performance', *ASHRAE Transactions*, vol. 128, no. Part2, pp. 378–386, Jul. 2022.

[114] D. Zhan, N. Sezer, D. Hou, L. Wang, and I. G. Hassan, 'Integrating Urban Heat Island Impact

into Building Energy Assessment in a Hot-Arid City', *Buildings*, vol. 13, no. 7, Art. no. 7, Jul. 2023, doi: 10.3390/buildings13071818.

[115] Y. Feng, Q. Duan, X. Chen, S. S. Yakkali, and J. Wang, 'Space cooling energy usage prediction based on utility data for residential buildings using machine learning methods', *Applied Energy*, vol. 291, p. 116814, Jun. 2021, doi: 10.1016/j.apenergy.2021.116814.

[116] Z. Wang, T. Hong, and M. A. Piette, 'Building thermal load prediction through shallow machine learning and deep learning', *Applied Energy*, vol. 263, p. 114683, Apr. 2020, doi: 10.1016/j.apenergy.2020.114683.

[117] U. Ali *et al.*, 'Urban building energy performance prediction and retrofit analysis using data-driven machine learning approach', *Energy and Buildings*, vol. 303, p. 113768, Jan. 2024, doi: 10.1016/j.enbuild.2023.113768.

[118] A. G. ASHRAE, 'Guideline 14-2014: measurement of energy, demand, and water savings', *American society of heating, refrigerating, and air conditioning engineers, Atlanta, Georgia*, 2014.

[119] L. Wang, R. Kubichek, and X. Zhou, 'Adaptive learning based data-driven models for predicting hourly building energy use', *Energy and Buildings*, vol. 159, pp. 454–461, Jan. 2018, doi: 10.1016/j.enbuild.2017.10.054.

[120] G. Hong, G.-S. Choi, J.-Y. Eum, H. S. Lee, and D. D. Kim, 'The Hourly Energy Consumption Prediction by KNN for Buildings in Community Buildings', *Buildings*, vol. 12, no. 10, Art. no. 10, Oct. 2022, doi: 10.3390/buildings12101636.

[121] Z. Dong, J. Liu, B. Liu, K. Li, and X. Li, 'Hourly energy consumption prediction of an office building based on ensemble learning and energy consumption pattern classification', *Energy and Buildings*, vol. 241, p. 110929, Jun. 2021, doi: 10.1016/j.enbuild.2021.110929.

[122] R. Mathumitha, P. Rathika, and K. Manimala, 'Intelligent deep learning techniques for energy consumption forecasting in smart buildings: a review', *Artif Intell Rev*, vol. 57, no. 2, p. 35, Feb. 2024, doi: 10.1007/s10462-023-10660-8.

[123] A. J. Guízar Dena, M. del C. García Hípola, and C. Fernández Bandera, 'Optimization testing for the modeling and characterization of three-dimensional elements to enhance interoperability from building information modeling (BIM) to building energy modeling (BEM)', *Energy and Buildings*, vol. 317, p. 114394, Aug. 2024, doi: 10.1016/j.enbuild.2024.114394.

[124] R. Benfer and J. Müller, 'Semantic digital twin creation of building systems through time series

based metadata inference – A review', *Energy and Buildings*, vol. 321, p. 114637, Oct. 2024, doi: 10.1016/j.enbuild.2024.114637.

[125] F. Vahdatikhaki, N. Salimzadeh, and A. Hammad, 'Optimization of PV modules layout on high-rise building skins using a BIM-based generative design approach', *Energy and Buildings*, vol. 258, p. 111787, Mar. 2022, doi: 10.1016/j.enbuild.2021.111787.

[126] G. Di Giovanni, M. Rotilio, L. Giusti, and M. Ehtsham, 'Exploiting building information modeling and machine learning for optimizing rooftop photovoltaic systems', *Energy and Buildings*, vol. 313, p. 114250, Jun. 2024, doi: 10.1016/j.enbuild.2024.114250.

[127] J. An and T. Hong, 'Energy harvesting using rooftops in urban areas: Estimating the electricity generation potential of PV and wind turbines considering the surrounding environment', *Energy and Buildings*, vol. 323, p. 114807, Nov. 2024, doi: 10.1016/j.enbuild.2024.114807.

[128] T. Hong, Y. Xu, K. Sun, W. Zhang, X. Luo, and B. Hooper, 'Urban microclimate and its impact on building performance: A case study of San Francisco', *Urban Climate*, vol. 38, p. 100871, Jul. 2021, doi: 10.1016/j.uclim.2021.100871.

[129] T. Hong, Y. Chen, X. Luo, N. Luo, and S. H. Lee, 'Ten questions on urban building energy modeling', *Building and Environment*, vol. 168, p. 106508, Jan. 2020, doi: 10.1016/j.buildenv.2019.106508.

[130] B. Bueno, L. Norford, J. Hidalgo, and G. Pigeon, 'The urban weather generator', *Journal of Building Performance Simulation*, vol. 6, no. 4, pp. 269–281, Jul. 2013, doi: 10.1080/19401493.2012.718797.

[131] M. Mortezazadeh and L. L. Wang, 'A high-order backward forward sweep interpolating algorithm for semi-Lagrangian method', *International Journal for Numerical Methods in Fluids*, vol. 84, no. 10, pp. 584–597, 2017.

[132] M. Mortezazadeh and L. Wang, 'An adaptive time-stepping semi-Lagrangian method for incompressible flows', *Numerical Heat Transfer, Part B: Fundamentals*, vol. 75, no. 1, pp. 1–18, 2019.

[133] M. Mortezazadeh and L. L. Wang, 'Solving city and building microclimates by fast fluid dynamics with large timesteps and coarse meshes', *Building and Environment*, vol. 179, p. 106955, 2020.

[134] T. Hong, Y. Chen, S. H. Lee, and M. A. Piette, 'CityBES : A Web-based Platform to Support City-Scale Building Energy Efficiency', in *Urban Computing 2016*, San Francisco, San

Francisco, California USA, 2016.

[135] M. Ferrando, F. Causone, T. Hong, and Y. Chen, ‘Urban building energy modeling (UBEM) tools: A state-of-the-art review of bottom-up physics-based approaches’, *Sustainable Cities and Society*, p. 102408, 2020.

[136] T. Hong, Y. Chen, M. A. Piette, and X. Luo, ‘Modeling City Building Stock for Large-Scale Energy Efficiency Improvements using CityBES Overview of CityBES’, no. Nyserda, 2016.

[137] Y. Chen, T. Hong, and M. A. Piette, ‘Automatic generation and simulation of urban building energy models based on city datasets for city-scale building retrofit analysis’, *Applied Energy*, vol. 205, pp. 323–335, 2017.

[138] C. Potter and O. Alexander, ‘Impacts of the San Francisco Bay Area shelter-in-place during the COVID-19 pandemic on urban heat fluxes’, *Urban Clim*, vol. 37, p. 100828, May 2021, doi: 10.1016/j.uclim.2021.100828.

[139] M. Peterson, ‘Extreme Heat Killed 14 People in the Bay Area Last Year. 11 Takeaways From Our Investigation | KQED’. Accessed: Apr. 24, 2024. [Online]. Available: https://www.kqed.org/science/1932903/extreme-heat-killed-14-people-in-the-bay-area-last-year-10-takeaways-from-our-investigation

Appendix A

To further elucidate the DVIndex, which serves as a measure of cluster compactness, we define the components contributing to it. The Intra(k) term represents the average sum of distances between cluster centers and individual data points (Eq. (A-1)). This Intra value is then normalized by dividing it by its maximum value (Eq. (4-3)), computed from Intra values obtained for different cluster numbers, ranging from k =4 to an upper limit defined as k (Eq. (A-2)). Separateness is the ratio of the maximum and minimum Squared Euclidean Distance (SED) between cluster centers, multiplied by the sum of inverse distances between cluster centers (Eq. (A-3)).

The IntraRatio and InterRatio terms collectively denote the overall compactness and separateness of clusters, respectively. These ratios normalized for comparison within the range of 0 to 1, convey insights into the clustering structure. Typically, the intra-term diminishes with increasing cluster numbers, indicating enhanced compactness, while the inter-term tends to rise with greater cluster counts, signifying increased separation between clusters. Consequently, the DVIndex becomes more meaningful when clusters are either merged or split, as the inter-term is particularly sensitive to changes in inter-cluster distances.

A modulating parameter γ is introduced to balance the relative importance of IntraRatio and InterRatio terms. Its value influences the sensitivity of the DVIndex to noise within the data. In scenarios where raw data exhibit minimal noise, γ is often set to 1. Conversely, if noise is present, γ is adjusted to a value less than 1 to mitigate its impact. Conversely, when prioritizing cluster separateness over compactness, γ is set to a value greater than 1.

$$Intra(k) = \frac{1}{N}\sum_{i=1}^{k}\sum_{x \in c_i}\|x - z_i\|^2 \qquad \text{(A-1)}$$

$$MaxIntra = \max_{k=4,\ldots,K}\left(Intra(k)\right) \tag{A-2}$$

$$Inter(k) = \frac{Max_{i,j}\left(\left\|z_i - z_j\right\|^2\right)}{Min_{i\neq j}\left(\left\|z_i - z_j\right\|^2\right)}\sum_{i=1}^{k}\left(\frac{1}{\sum_{x\in c_i}\|x - z_i\|^2}\right) \tag{A-3}$$

$$MaxInter = \max_{k=4,\ldots,K}\left(Inter(k)\right) \tag{A-4}$$

where *N* is the number of data (objects), *N*=365 in this study; *k* starts from 4, and z_i denotes the center of the cluster C_i. *i* = 1, 2,…, *k*-1; *j* = *i*+1, *i*+2, …, *k*.

Appendix B

In this study, we included six meteorological variables (outdoor air dry-bulb temperature, outdoor air relative humidity, wind speed, wind direction, normal solar irradiation, and diffuse solar irradiation) and ten building features (wall U value, roof U value, window U value, window SHGC, occupancy density, appliance power density, lighting power density, ventilation rate, heating set point, and cooling set point). We generated synthetic data using a physics-based building energy model to train and test the machine learning model, applying latin hypercube sampling (LHS) to perform parametric simulations.

The total number of data points was calculated as follows:

$$Total\ number\ of\ data\ points = N_{Sim} \times N_{hours} \times \left(N_{inputs} + N_{outputs}\right) \quad \text{(B-1)}$$

$$N_{Sim} = S \times F \quad \text{(B-2)}$$

where:

- N_Sim is the number of building energy simulations, here are EnergyPlus simulations.
- N_hours is the total number of hours in results, here is 8760 hours.
- N_inputs represents the number of input variables (16 in this study, 10 building features and 6 meteorological variables.)
- N_outputs is the number of output variables, we have one output variable in this work, hourly energy use intensity.
- S is sampling size, set to 50 to ensure that the sample distribution closely represents the original distribution.
- F is the number of building features, 10 in this study.

Using these parameters, we calculated approximately 82 million individual data points.