Urban Building Energy Models for District Cooling: A Data-Driven Approach Considering Building and Occupant Behavior Dynamics

Omar Ahmed

A Thesis
In the Department of
Building, Civil, and Environmental Engineering

Presented in Partial Fulfillment of the Requirements
for the Degree of
Master of Applied Science (Building Engineering)

At Concordia University
Montreal, Quebec, Canada

September 2024

CONCORDIA UNIVERSITY

SCHOOL OF GRADUATE STUDIES

This is to certify that the thesis prepared

By: Omar Ahmed

Entitled: Urban Building Energy Models for District Cooling: A Data-Driven Approach Considering Building and Occupant Behavior Dynamics

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

Master of Applied Science (Building Engineering)

Complies with the regulations of the university and meets the accepted standards with respect to originality and quality.

Signed by the final examining committee

______________________________ Chair

Dr. Bruno Lee

______________________________ Examiner

Dr. Mohamed Ouf

______________________________ Supervisor

Dr. Liangzhu (Leon) Wang

______________________________ Supervisor

Dr. Ibrahim Galal Hassan

Approved by: ______________________________

Chair of Department or Graduate Program Director: Dr. S. Samuel Li

______________ 2024 ______________________________

Dean of Faculty: Dr. Mourad Debbabi

ABSTRACT

Urban Building Energy Models for District Cooling: A Data-Driven Approach Considering Building and Occupant Behavior Dynamics

Omar Ahmed

District cooling offers an energy-efficient solution for hot urban regions where cooling demands are high. Accurate and rapid predictions of cooling requirements are vital during the planning phase to support informed decision-making. Surrogate models, which combine physics-based simulations with statistical or machine learning techniques, can harness the strengths of both methods, leading to more accurate building energy predictions at a low computational cost. In this study, a surrogate model, which combines machine learning with building physics-based archetypes, is employed to predict the cooling energy use intensities for high-rise buildings in a mixed-use district. The proposed surrogate models predict the impact of building design parameters, building operation characteristics, and occupant-related parameters on building energy performance. High-rise building models, representative of the district, are created using EnergyPlus software. The detailed cooling load profiles of these baseline models are simulated, analyzed, and validated against measured data and literature benchmarks. The resulting cooling loads are then aggregated at the district level, providing a physics-based method for urban-scale energy prediction. Parametric simulations are automated in RStudio using the developed archetypes by altering key parameters such as building envelope characteristics, geometry, and operational parameters, including occupant behavior. The resulting datasets are used to train machine learning models to approximate the outcomes of physics-based simulations. Additionally, the trained models are integrated into a user-friendly interface, enabling computationally efficient predictions of cooling requirements for each building in the district. The developed models show excellent performance, with R^2 values near 1 and RMSE below 0.17 kWh/m²/month on unseen data. This study demonstrates the potential of surrogate machine learning in predicting and optimizing building energy performance under different design, operation, and occupancy settings. It also provides insights into the impact of training dataset size on the accuracy of surrogate machine learning models.

Key Words: Urban Building Energy Modeling, District Cooling, Surrogate Modeling, Machine Learning, Building Performance Analysis, Occupant Behavior.

ACKNOWLEDGEMENTS

I would like to express my sincere appreciation for the funding provided by the NPRP award [#NPRP11S-1208-170073] from the Qatar National Research Fund (a member of The Qatar Foundation). This research would not have been possible without their generous support. I am deeply grateful to my supervisors, Dr. Liangzhu (Leon) Wang and Dr. Ibrahim Galal Hassan. Dr. Wang's patience, guidance, and unwavering commitment as my daily supervisor have been invaluable. He consistently sacrificed his time to discuss my research and provided support in every way possible. I am also immensely thankful to Dr. Ibrahim Galal Hassan for his guidance, patience, mentorship, and the significant time he invested in helping me with this thesis and other publications. Under his mentorship, I learned to approach problems from a broader perspective and to manage my time and efforts effectively. I have been profoundly inspired by the enthusiasm and dedication both of my supervisors demonstrate in their scientific research and daily work. I would also like to express my appreciation for the friends I was fortunate to make throughout this journey. Finally, and most importantly, I am deeply thankful for the patience, love, and kindness my family has shown me while being away from home.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

NOMENCLATURE

AI	Artificial Intelligence
ANN	Artificial Neural Network
IEA	International Energy Agency
ASHRAE	American Society of Heating Refrigeration and Air Conditioning Engineers
BEM	Building Energy Model
CART	Classification and Regression Trees
CV	Coefficient of Variance
DC	District Cooling
ECM	Energy Conservation Measures
EUI	Energy Use Intensity
GBRT	Gradient Boosting Regression Trees
GCC	Gulf Cooperation Council
GFA	Gross Floor Area
GSAS	Gulf Sustainability Assessment System
HVAC	Heating Ventilation and Air Conditioning
LHS	Latin Hypercube Sampling
LR	Linear Regression
ML	Machine Learning
MLP	Multi-Layer Perceptron
MLR	Multivariate Linear Regression
NRMSE	Normalized Root Mean Squared Error
OB	Occupant Behavior
PDF	Probability Density Function
PNNL	Pacific Northwest National Laboratory
$\mathbf{R^2}$	R square
RF	Random Forest
RMSE	Root Mean Square Error
SHGC	Solar Heat Gain Coefficient
SSE	Sum of Squared Error
SVM	Support Vector Machine
TMY	Typical Meteorological Year
TR	Ton Refrigeration
UAE	United Arab Emirates
UBEM	Urban Building Energy Modeling
WWR	Window-to-wall ratio
XGB	Extreme Gradient Boosting

CHAPTER 1

1. Introduction

1.1. Problem Statement

As per the International Energy Agency (IEA), more than 30% of global energy consumption and 55% of the overall electricity demand stem from the building sector [1]. The predominant contributors to energy use in this sector are heating and cooling, accounting for over half of the total energy consumption. With the ongoing trend of urbanization and the continual growth of urban populations, there is an anticipated further increase in energy consumption within the building sector. The utilization of building energy modeling is crucial for analyzing building performance, identifying potential energy-saving measures, and assessing the impact of diverse design and operational strategies on the energy efficiency of buildings, ultimately aiming to optimize their design and operation. In the realm of building energy modeling, it is imperative to account for the distinctive climatic conditions of a particular region. Achieving optimal energy performance necessitates a comprehensive examination of buildings, taking into consideration factors such as their physical footprint and the prevailing climate zone. For example, buildings situated in cold and arid climate zones demand different design and operational approaches compared to those in hot and humid climates.

In areas characterized by hot and arid climates such as Qatar, specifically those falling under the Koeppen-Geiger classification of BWh climate, which refers to regions having hot and dry desert climates with an annual average temperature exceeding 18 ° C [2], a substantial portion of the building sector's electricity consumption is allocated to cooling functions. For Qatar, an in-depth analysis of the energy distribution profile revealed that the residential and commercial sectors contributed 41% and 19%, respectively, to the overall electricity consumption [3]. Notably, 65 % of the electricity used in Qatar is used for cooling systems, with traditional air conditioning, being the predominant type [4]. Considering the ongoing trends of population growth, urbanization, and the evolving climate due to global warming, there is an anticipated growth in cooling requirements. Consequently, the development and implementation of cooling systems and technologies present substantial opportunities for energy conservation and hold valuable potential for mitigating the growth in electricity consumption and peak demand.

In the Gulf Cooperation Council (GCC) region, District Cooling (DC) has emerged as a viable and recent cooling approach aimed at reducing overall energy consumption in buildings and mitigating associated environmental pollution [4]. Research indicates that district cooling systems offer the adoption of cost-effective and energy-efficient solutions, demonstrating enhanced operational flexibility and utilization of low-cost energy storage options. Furthermore, they leverage energy sources such as excess heat from industries and natural cooling that might otherwise remain untapped [5]. Owing to these advantages, Qatar has experienced substantial growth in installed DC plant cooling capacity, witnessing an almost two-fold increase from 562,000 TR in 2016 to 995,700 TR in 2020 [6].

The precise estimation of the cooling load holds paramount importance in the design of a district cooling system as it directly influences the system's design, operation, and cost-effectiveness. However, accurately estimating the cooling load characteristics of a DCP remains a formidable challenge. Numerous building developers tend to overestimate the cooling consumption of buildings, leading to the inefficient utilization of district cooling plant capacity. This practice not only results in wasted resources but also has a direct impact on the comfort index within the conditioned indoor environment. To support decision-making for stakeholders in district cooling plants, there is a crucial need for accurate and quick prediction and analysis of building and district-level cooling energy requirements.

Urban Building Energy Models (UBEMs) prove to be promising tools for predicting operating energy consumption and indoor conditions for clusters of buildings within large-scale simulations in diverse urban scenarios [7]. Building energy consumption prediction can generally be categorized into three groups. First, there is the building physical energy model, commonly referred to as the "white box," which is based on precise building characteristics and heat and mass balance formulas. Secondly, the data-driven model, also referred to as the "black box," is based on machine learning algorithms and building historical data. Finally, the hybrid model, sometimes known as the "grey box," is a methodology that combines easily accessible data with building physical information [8].

Various studies have consistently demonstrated that grey-box models effectively strike a balance, harnessing the strengths inherent in both white-box and black-box models, and yielding superior prediction results [9]. However, their widespread adoption is limited due to ambiguity in their development methods and the lack of widely used development software for grey-box models [10]. Regardless of the methodology followed when creating UBEMs, uncertainties stemming from various sources, including assumptions about occupant behavior (OB), may constrain their reliability and potential [11]. Despite OB being acknowledged as a main factor influencing building energy consumption, it is still a relatively unexplored topic in the field of UBEM. Most UBEMs rely on deterministic schedules and parameters for defining OB, with only a few adopting a probabilistic evaluation [12]. Studies have demonstrated that utilizing improper OB models can lead to oversizing district energy systems, which would increase investment and decrease their operational efficiency [11]. While relying on constant occupancy profiles is not recommended for applications at the urban scale, obtaining dependable information on building occupancy at an urban scale poses a challenge due to the cost and complexity associated with deploying numerous sensing devices across the entire urban area [13].

1.2. Research Objectives

To address the two problems discussed in section 1.1, the research has two main objectives:

(1) To estimate the energy consumption of a mixed-use district, a surrogate machine learning model has been developed to predict the cooling energy use intensities for the three available building types (Residential, Commercial, and Mixed-Use). A hybrid method, combining representative building physics-based models and machine learning techniques is used. Twelve typical high-rise building archetypes are modeled to represent each building type within the case study area. The cooling load profiles of the baseline archetype models are simulated, analyzed, and validated. Following this validation, parametric simulations are conducted. For each building type, a total of 7,800 cases are generated by varying critical parameters related to the building envelope, geometry, and occupant behavior. Three neural network models, one for each building type, are then developed using the created input-output datasets for cooling energy use intensity predictions.

(2) A user-friendly decision support environment is created by linking the trained machine-learning models to an Excel sheet. This tool is used to predict the monthly cooling energy use intensity profile for each building in the case study district using easily definable inputs. The results from these individual predictions are then aggregated to obtain the district-level cooling energy consumption profile. By including inputs related to occupant behavior, the tool can more accurately represent the diversity between buildings of the same type.

CHAPTER 2

2. Literature review

2.1. Overview

This literature review explores the current state of Urban Building Energy Modeling (UBEM) and its relationship with District Cooling Design. It starts by discussing District Cooling Systems (DCS) and their advantages in densely populated urban areas. The scope of UBEM, covering everything from urban planning to carbon reduction strategies, is outlined. UBEM Top-down and Bottom-up Methodologies are discussed, with a closer look at the Bottom-up UBEM approaches. Physics-based Bottom-up UBEM, leveraging heat and mass balance equations, Data-driven Bottom-up UBEM, involving statistical and Artificial Intelligence (AI) methods, and Hybrid Bottom-up UBEM models, combining simulations with machine learning or statistical approaches, are all elucidated, to put forward the rationale for adopting a hybrid modeling approach. To move closer to the research gap, occupant behavior is then identified as a major contributor to building energy usage before outlining the importance of its representation in building energy analysis and what limits its depiction in urban scale energy analysis. Finally, this section ends with pointing out the potential contributions from this work to areas where more research is needed in this interdisciplinary field.

2.2. District Cooling Design

Currently, the majority of global building cooling demands are satisfied by traditional on-site cooling systems. Cold energy is generated and distributed at the end user's location in such systems, whether they are smaller capacity window units or central air- or water-cooled chillers for larger applications [14]. A District Cooling System (DCS) is characterized as a system that provides end users with chilled water from a central source for the purposes of dehumidification and space cooling [15]. Generally, it comprises four key components: the heat rejection system, the central chiller plant, the end users, and the distribution system which links the previous three together. For dense districts experiencing hot climatic conditions throughout the year, rapid urbanization, and extensive building developments, such as those found in the Gulf Cooperation Council (GCC) regions, District Cooling (DC) systems present several significant advantages over conventional cooling systems. DC requires less energy, can be integrated with several renewable and sustainable energy technologies, facilitates a more flexible system operation, offers affordable energy storage solutions, and enables the use of energy sources such as excess heat from industries and natural cooling that would otherwise go unused [5], [16], [17]. The cooling load is usually considered the most significant factor influencing the design, performance, and decision-making process related to DC systems. Therefore, a detailed analysis of the cooling requirements and usage patterns is required since they can vary significantly for diverse end users.

Quantifying cooling loads is a preliminary step in the DC design and analysis process. Both peak cooling load requirements and annual cooling load profiles are needed for the design and analysis. Engineering standards such as ASHRAE provide a straightforward method for predicting district-level cooling loads. Comprehensive tables for cooling load density data per unit area are provided for different building types. However, given their limits in terms of accuracy, these data should be used cautiously in the initial stages of DC planning. At the district level, many research papers have addressed more sophisticated cooling load calculation and analysis techniques. Actual measured end-user data is used in some approaches for cooling load calculation [18], [19]. However, these methods are only applicable to existing buildings, and even then, can often be missing or unavailable. For districts still in the design phase, simplified engineering approaches or building performance simulation tools are commonly used for load estimation [20], [21]. To further clarify the scope of this work, the state-of-the-art in UBEM is presented in the following section.

2.3. Urban Building Energy Modeling

Urban Building Energy Modeling involves the development and analysis of a dependable building energy model for a cluster of buildings within an urban setting [22]. UBEM operates within a spatial scale ranging from a city block to a district and extends to an entire city. The focus extends beyond individual buildings to encompass the interactions between buildings and the impact of the urban microclimate [23]. UBEM serves as a valuable tool for aiding the design and optimization of urban buildings on a comprehensive scale, aligning with goals related to energy efficiency, sustainability, and resilience within urban landscapes [22]. Four main categories have been proposed for classifying UBEM applications [24] (Figure 1).

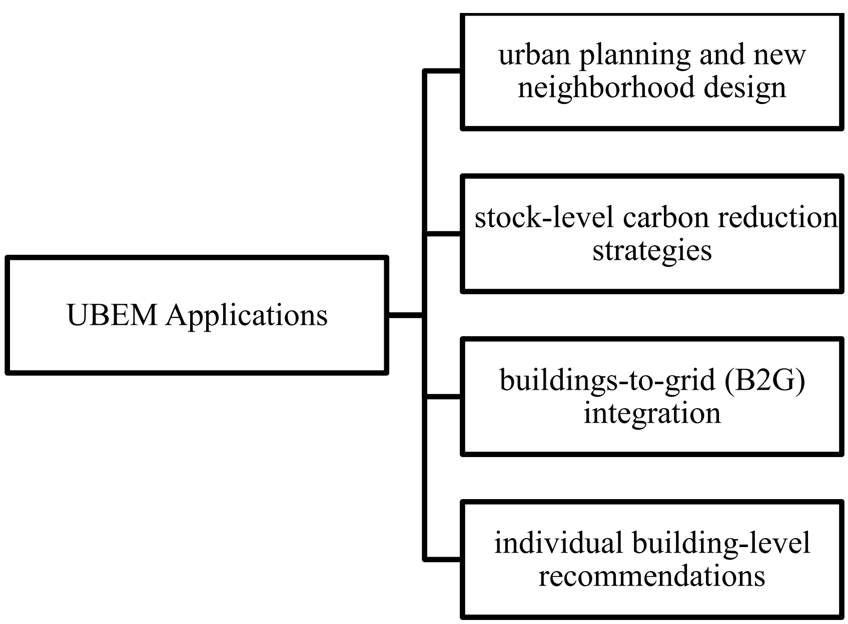


Figure 1. UBEM Applications [24]

1. Urban planning and new neighborhood design.

Urban characteristics such as geometry, typology, shading, daylight, and urban heat island effects can significantly affect building energy use [25], [26]. In the context of designing new neighborhoods and urban planning, UBEMs are particularly valuable. An UBEM can be used in the preliminary design stage to explore various building massing layouts, window-to-wall ratio (WWR) configurations, etc., This comparative analysis helps gain insights into the relative performance of different planning scenarios. [27], [28]

2. Stock-level carbon reduction strategies.

For stock level energy and carbon reduction strategies, the potential energy savings resulting from specific upgrades when applied universally to buildings with similar characteristics, such as program type, age, category, or archetype in a district need to be comprehended. For this purpose, UBEMs are used for estimating energy savings by replacing existing templates with new ones that incorporate the upgrades. Through modeling and simulation, the results can reveal the overall potential for carbon reduction and energy savings across different retrofit scenarios, pinpointing the building types that contribute most significantly to these savings. [29], [30]

3. Individual building-level recommendations.

As opposed to municipal governments and policymakers who are primarily concerned with stock-level analyses, building owners are more focused on understanding the specific energy savings resulting from particular upgrades to their buildings or portfolios, to establish customized recommendations. To address this objective, UBEMs calibrated to the individual building level, or auto-calibrated building-level UBEMs can predict energy savings for individual buildings within a specific region based on metered annual or monthly data for the buildings under study. [31], [32], [33]

4. Buildings-to-Grid (B2G) integration

Due to the substantial impact of buildings on the grid and the advantages of data-driven models for predictive and demand-responsive controls, demand response has now become a viable option for everyday application [34]. Buildings can function as both controllable consumers and producers of energy, presenting unique opportunities for building-to-grid (B2G) integration. UBEMs, which generate current and future hourly load profiles for buildings, appear to be an ideal complement to supply-side modules, enabling integrated analysis. Electrification scenarios or load control strategies based on pricing could be further examined using more integrated grid -building models to assess potential new chokepoints in the grid as the overall load increases during operations. B2G integration emerges as a solution to coordinate meeting this load without significantly altering the capacity of the existing power system. [35], [36], [37]

The modeling procedure in UBEM is not a straightforward process and involves various challenges and uncertainties. Selecting the appropriate modeling procedure is highly dependent on the required objective and application, the scale of the model, and available data and resources. A number of classifications for UBEM approaches have been found in the literature, with a more recent and widely cited classification presented in [38] (Figure 2).

2.3.1. Top-down UBEM

The Top-down modeling approach is typically data-driven, employing statistical and regression models to explore the relationships between urban energy usage patterns and associated drivers such as macro and social economics as well as energy policies [39]. Because of their straightforward modeling approach, relatively high processing speeds, and the ready availability of necessary data, top-down models have found extensive use in urban building energy-related studies [40], [41]. However, these models are not considered suitable for in-depth analysis and building-level recommendations since they regard a collection of buildings as a unified energy asset [39]. Furthermore, their reliance on aggregated historical data without precise spatial or temporal information hinders the prediction of future trends [42] and limits their ability to examine technological changes in current and future scenarios [12].

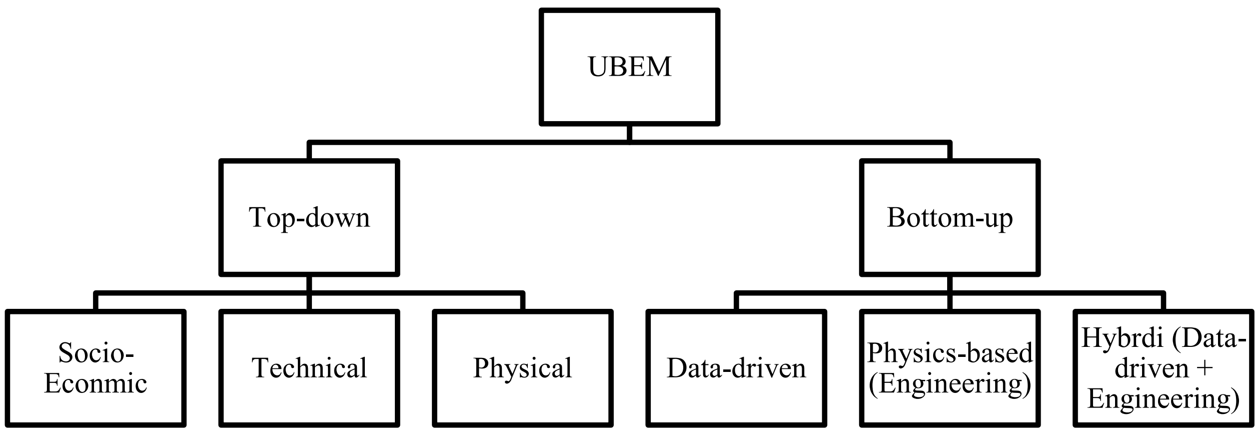


Figure 2. UBEM approach classification [38]

2.3.2. Bottom-up UBEM

In contrast, bottom-up models are characterized by analyzing the energy demand at the individual building level and consequently scaling up to the urban level. While this approach demands extensive disaggregated data and computational resources, it offers higher spatiotemporal resolution and model accuracy, which can support the decision-making processes in energy conservation measures and possible future scenarios [39]. Bottom-up models are also able to consider the dynamic interactions between buildings and the urban setting, allowing the inclusion of more specific building and urban environmental aspects in the model. Bottom-up models can be classified into three categories, Physics-based (Engineering), Statistical and data-driven, and Hybrid approaches.

2.3.2.1 Physics-based (Engineering) approaches

Physics-based models leverage a building's unique technological and physical characteristics, using heat and mass balance equations, in addition to information from heating and cooling systems, weather, building features, and construction details to determine energy demand [42]. These models are highly adaptable for evaluating scenarios related to energy efficiency and technology improvements. The bottom-up archetypal method has been extensively employed in the urban setting to evaluate the impact of adopting new technologies and energy efficiency policies [22], [29]. To estimate the building stock energy consumption, this method entails modeling each building archetype in a simulation engine and then scaling up the estimations for the regional or national level [43]. Building physics-derived quantitative data is a major component of these approaches, requiring inputs such as internal and external temperatures, HVAC system characteristics, building components' thermal properties (U values), internal load definitions, occupancy, and building schedule information [23], [44]. Additionally, these models require a significant quantity of technical data to predict energy usage and multiple assumptions to describe occupant behavior [45]. Physics-based building energy modeling research has grown rapidly over the last few decades. CityBES [46], CitySim [47], City Energy Analyst (CEA) [48], Urban Modeling Interface (UMI) [49], Open Integrated District Energy Assessment by Simulation (OpenIDEAS) [50], and TEASER [51] are considered the most well-known readily available tools used for various applications in the urban energy domain.

2.3.2.2 Data-driven bottom-up UBEM

Data-driven models in the bottom-up approach can capture the relationships between the energy consumption of a building and the buildings' characteristics along with other drivers [9]. Existing data sources such as survey data, building stock statistics, billing data, and socioeconomic parameters are leveraged for predicting building energy consumption [44]. Data-driven modeling typically encompasses statistical and Artificial Intelligence (AI) (machine learning) approaches [52]. In the statistical approach, regression techniques are employed to establish inverse mathematical models based on building design or operational parameter details. Linear Regression (LR), Non-linear Regression (NR), Multiple Linear Regression (MLR), and Conditional Demand Analysis (CDR) are the most frequently adopted algorithms in this method [52]. Kontokosta [53] utilized MLR using the New York City Energy Benchmarking (2011) dataset to predict urban buildings' energy usage. Kuusela et al. [54] also applied MLR to predict energy consumption based on building characteristics at the neighborhood scale. Mastrucci et al. [55] adopted a graphical information system (GIS) based statistical approach to estimate the energy usage of residential urban areas. Using an autoregressive model, Dagnely et al. [56] evaluated the accuracy of linear regression using the ordinary least squares (OLS) and support vector machine (SVM) approaches, demonstrating that both approaches may deliver a performance accuracy level that is satisfactory. While the MLR algorithm is extensively utilized for predicting energy use due to its simplicity and interpretability, it faces limitations in capturing non-linear and complex patterns.

The AI approach primarily relies on machine learning (ML) techniques to predict urban building energy usage by automatically detecting and learning patterns in the data. To establish the mathematical relationship between building energy use and significant variables such as building features, urban characteristics, and occupancy features, the model is trained and learns using historical datasets [52]. ML techniques are generally categorized into two main groups: supervised and unsupervised learning. Supervised learning involves predicting the output by establishing a complex relationship between multiple features and includes regression and classification algorithms. Classification algorithms predict output data when the outcome is a label, such as building typologies or energy ratings, while regression algorithms gain knowledge from the input data and predict actual output values, such as energy consumption. On the other hand, unsupervised learning algorithms identify underlying structures, correlations, or unidentified patterns in the input data, such as energy usage patterns. Data-driven models are extremely useful in various urban energy applications, including prediction, forecasting, benchmarking, mapping, and classification [44]. Rahman et al. [57] applied deep recurrent neural networks to forecast medium- to long-term electricity usage for residential and commercial buildings. Using national survey data, Robinson et al. [58] suggested an approach for estimating the energy consumption of commercial buildings by training several machine learning models (gradient boosting model, linear regression, support vector regression, random forest). Zhang et al. [59] presented a data-driven approach that considers a number of variables, including building attributes, geometry, and urban morphology, to estimate energy consumption and greenhouse gas emissions. Abbasabadi et al. [60] proposed an integrated framework for urban energy use modeling which models energy used in urban buildings and transportation leveraging a ML approach. Razak et al. [61] created a ML model that uses building design features at the early development stages to predict yearly average energy use. Wurm et al. [62] created a workflow employing deep learning algorithms to simulate the building stock heating demand. To summarize, machine learning approaches can improve the accuracy of urban energy use forecasting by utilizing high temporal and spatial resolution data and sophisticated algorithms that facilitate the capture of complicated and non-linear patterns. The primary challenge lies in the insufficient availability of high-quality data available in large enough volumes to effectively train prediction models. This underscores the need for a robust building energy modeling approach that can reliably forecast the energy performance of whole building stocks, especially in the event that resources for intricate decision-making analysis are scarce.

2.3.2.3 Hybrid bottom-up UBEM

Hybrid models aim to mitigate the shortcomings of both data-driven and physics-based models, by combining various techniques to achieve the unique goals of each model. These models integrate elements of ML or statistical approaches with physics-based simulations to leverage the strengths of both approaches, incorporating factors that each model alone may not be able to capture comprehensively. In this approach, buildings can be modeled using their physical attributes, while stochastic parameters are represented using empirical distributions to account for their uncertainty [12]. The outputs from the physical simulations along with building metadata can then be used to predict building energy needs or other pertinent performance indicators. These hybrid approaches are often referred to as surrogate ML models. Naji et al. [63] developed three surrogate ML models (artificial neural network (ANN), genetic programming, and extreme learning machine (ELM)) for predicting the heating and cooling energy needs of a residential building in Turkey. Their findings emphasized the robustness and superior predictive performance of the ELM algorithm compared to ANN and genetic programming in predicting building energy consumption using the main building envelope properties. Melo et al. [64] developed surrogate models for estimating annual cooling loads for commercial buildings in Brazil employing five different algorithms (multiple linear regression (MLR), multivariate adaptive regression splines (MARS), support vector machines (SVM), ANNs, and Gaussian Processes (GP)), finding that ANNs outperformed the other algorithms on the base of normalized root mean squared error (NRMSE). Ascione et al. [65] generated two families of surrogate ANNs. One family assesses the energy performance of the existing building stock, while the other estimates the impact of energy retrofit measures. Their methodology, applied to Italian office buildings, aims to predict energy consumption for space conditioning and occupants' thermal comfort for any member of a building category in a reliable and computationally inexpensive manner. Magalhães et al. [66] developed surrogate ANNs utilizing simulation results from ESP-r to describe the relationship between heating energy use and indoor temperatures for European residential buildings. Papadopoulos et al. [67] employed tree-based ensemble methods, namely random forest (RF), gradient boosting regression trees (GBRT), and extremely randomized trees, along with a data set created using Ecotect, for developing surrogate models capable of predicting annual cooling and heating loads for residential buildings in Greece. Their findings emphasized the suitability of tree-based ensemble learning algorithms for building energy predictions, with GBRT outperforming RF and extremely randomized trees in their case study. Nutkiewicz et al. [32] proposed a hybrid methodology that combines simulation-based techniques employing the EnergyPlus tool with data-driven approaches utilizing a Convolutional Neural Network (CNN) model to achieve precise predictions of urban energy consumption across various temporal scales (hourly, daily, and monthly intervals), applied for 22 university buildings in California. Lopes et al. [68] developed surrogate ANNs to predict the annual cooling energy consumption of Brazilian office buildings. Additionally, they introduced and validated a new climate indicator designed to extend the applicability of their model across different climate zones. Jihad et al. [69] developed an ANN employing synthetic data derived from EnergyPlus simulations to predict heating and cooling loads for residential buildings in Morocco. Ngo [70] used simulation results from TRACE 700 to develop models for predicting office building cooling loads in Taiwan. The performance of individual ML algorithms (ANN, linear regression (LR), support vector regression (SVR), and classification and regression trees (CART)) was assessed, as well as strategic combinations of these ML models to create ensemble models. D’Amico et al. [71] developed a decision-support tool capable of analyzing the energy and environmental performance of Italian office buildings along their life cycle. Their tool was based on ANNs, developed using input-output data from TRNSYS simulations. Vázquez-Canteli et al. [72] developed a surrogate model for calculating the thermal losses and solar gains of urban buildings in the US leveraging simulation results from CitySim and two deep neural networks. Their surrogate model reduced the computational time by approximately 2500-fold compared to traditional urban scale simulation tools while maintaining a reasonable level of accuracy. Ciulla et al. [73] developed ANNs to predict the heating energy demand of non-residential European buildings. They created the necessary energy database for model training using simulations in the TRNSYS environment, adhering to European standards and regulations, and considered three different locations to represent various climatic conditions. Using meteorological data and building characteristics, Westermann et al. [74] devised a model that spans arbitrarily many locations for estimating heating and cooling demand for an office building in Canada with a bias of less than 3%. This model was developed by employing output data from EnergyPlus simulations alongside a deep temporal convolutional neural network. Lou et al. [75] leveraged TRNSYS simulation results to develop three ML models (ANN, SVM, and long-short-term-memory neural network) for simultaneous prediction of heating, cooling, lighting, and building integrated photovoltaic (BIPV) loads for an office building in the UK. Mui et al. [76] created a hybrid model to forecast the cooling energy consumption of residential buildings in Hong Kong by combining EnergyPlus and Artificial Neural Networks (ANN). Their hybrid model was then employed to assess how different building materials, construction techniques, and indoor-outdoor temperature fluctuations affect cooling energy consumption in a computationally inexpensive manner. Elbeltagi et al. [77] developed an ANN capable of predicting the energy use intensity (EUI) of Egyptian residential buildings using results from EnergyPlus simulations. Li et al. [78] trained ten different ML models for predicting space cooling and heating EUIs for both residential and non-residential buildings in China using data generated via the Urban Modelling Interface (UMI) physics-based tool. Their findings demonstrate that surrogate ML models can accurately predict building heating and cooling energy at multiple scales. Lee et al. [79] utilized EnergyPlus results to develop ANNs for predicting heating loads in residential buildings in Korea. They trained multiple ANNs using various combinations of input variables, and assessed the accuracy of the models, emphasizing the significance of selecting appropriate input variables for effective ANN model training. Liu et al. [80] developed three different surrogate ML models (RF, SVM, ANN) for predicting the energy consumption of a university building in northern China. Their approach aimed to optimize building envelope design parameters, demonstrating the advantages of the RF model in their case study. Jia et al. [81] evaluated the performance of four ML algorithms (MLR, SVM, extreme gradient boosting (XGB), and ANN) in predicting the monthly cooling EUIs of high-rise residential buildings in Qatar. The models were developed using synthesized data from EnergyPlus simulations, with the ANN demonstrating superior prediction accuracy compared to the other algorithms. Santos-Herrero et al. [82] developed surrogate ANNs to forecast operative temperatures and energy consumption across various spaces within an office building situated in Spain. Ali et al. [38] evaluated the performance of various ML algorithms (LR, RF, gradient boosting (GB), XGB, and adaptive boosting (AdaBoost)) as surrogates to a BPS model in predicting building performance metrics across different operational scenarios. Their results highlight that while XGB outperformed the other algorithms in predictive accuracy, showing the highest coefficient of determination value, the LR model demonstrated quick training times and straightforward interpretability while maintaining competitive prediction accuracies.

2.4. Occupant behavior modeling in building performance simulations

Occupant Behavior (OB) is one of the six major drivers of building energy consumption as shown in Fig. 3 [83]. OB can mainly be classified into occupancy and occupants' interaction with building systems (e.g., HVAC, lighting, appliances, and other energy equipment), as presented in Fig. 4 [84]. OB defines the presence and movement of occupants as well as their interaction with building systems, which are conventionally represented by fixed schedules in BPS tools. However, these static schedules are incapable of representing the actual impact of occupant presence on building energy consumption and the dynamic relationship between a building and its' occupants [85], [86], [87].

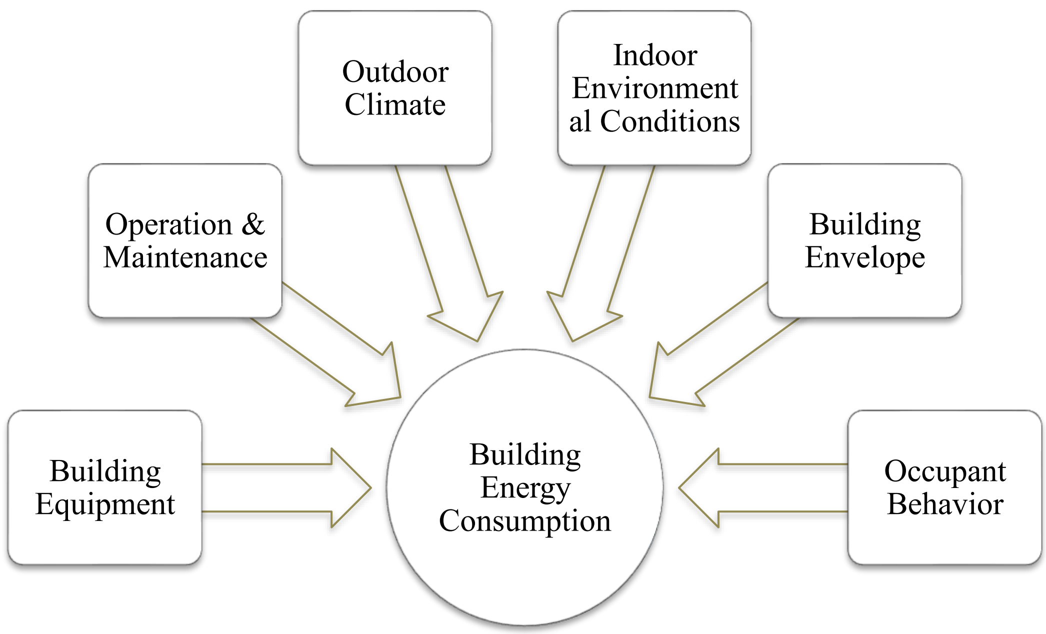


Figure 3. The six major drivers of building energy consumption.

Although five of the six major driving factors of building energy use mentioned in Fig. 3 have witnessed notable progress in their accurate representation in BPS tools, it remains a challenge to represent OB due to its unique characteristics [88], [89], [90], [91]. First, OB is stochastic; occupants do not consistently repeat the exact behavior daily since their behavior is governed by various factors. Second, OB is diverse as occupants have varying comfort demands and tolerances, which in turn causes occupants to exhibit a variety of behaviors for similar triggers. Third, OB is complex as it can be impacted by many components. Finally, OB has a dynamic nature which entails that OB is dependent on the buildings' design. For example, without analyzing the use of window coverings by occupants, simulation results could suggest that increasing the area of the window would result in increased daylight usage. However, due to glare issues, overly sized windows may just compel people to close curtains and rely only on electric lighting [88], [92], [93].

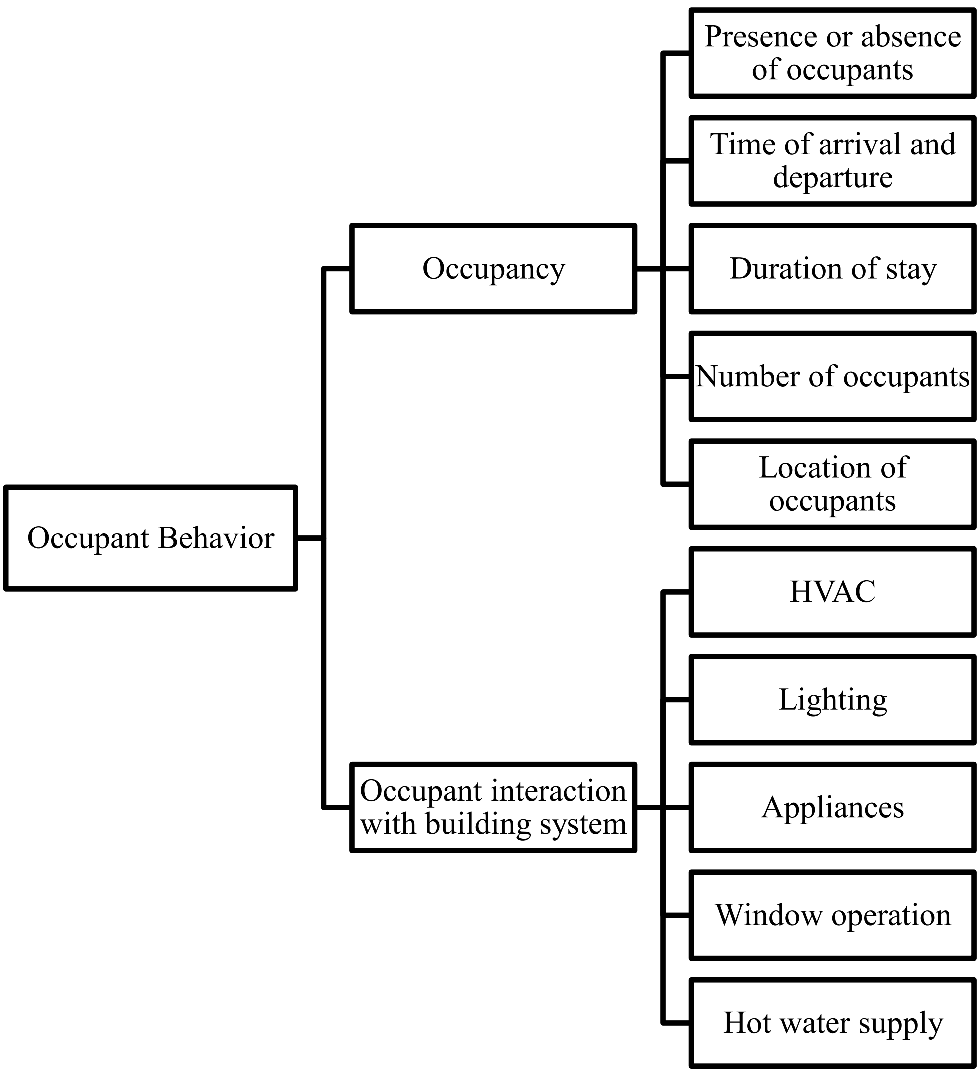


Figure 4. The two main categories of OB.

Due to the abovementioned characteristics of OB, several research groups studied how the uncertainty of input parameters in an OB model affects the building energy use [94], [95], [96], [97], [98]. For instance, a case study by Eguaras-Martínez et al. [99] demonstrated that the use of default occupancy schedules and the use of more realistic schedules created from actual collected data can cause differences of up to 30% for the heating and cooling demands of buildings. Another study by Li et al. [100] analyzed the electricity consumption for air-conditioning in 25 different apartments of a large residential building in Beijing in the summer. Although the building envelope was identical, the measured electricity consumption varied broadly up to a factor of ten among different apartments. This large variation was attributed to the air-conditioning system's operating mode. Apartments in which the occupants operated the AC in larger areas or for longer periods exhibited a higher energy consumption. The study concluded that the occupants' actions, not the apartments' design, were the main drivers of energy consumption.

The development and implementation of OB models in BPS tools require a number of steps which are elucidated in Figure 5 [101]. Before developing an OB model, it is important to precisely characterize the problem or issue that it is meant to address. The appropriate level of modeling and the models' balance between accuracy and applicability can then be specified based on the definition of this problem. For example, rather than simulating the stochastic behavior of each occupant individually when attempting to quantify a specific level of energy use in a large building, a simple model of the building as a whole would be sufficient because the stochastic nature of the occupants can be accounted for by statistically aggregating the energy use of a number of spaces [102]. On the other hand, when modeling occupants in a single zone, considering the stochastic nature of occupants would be more important [88]. It is well known that not all the components that influence OB can be accounted for and represented in an OB model, but models are generally created with the purpose of offering a decent estimate of OB in most cases.

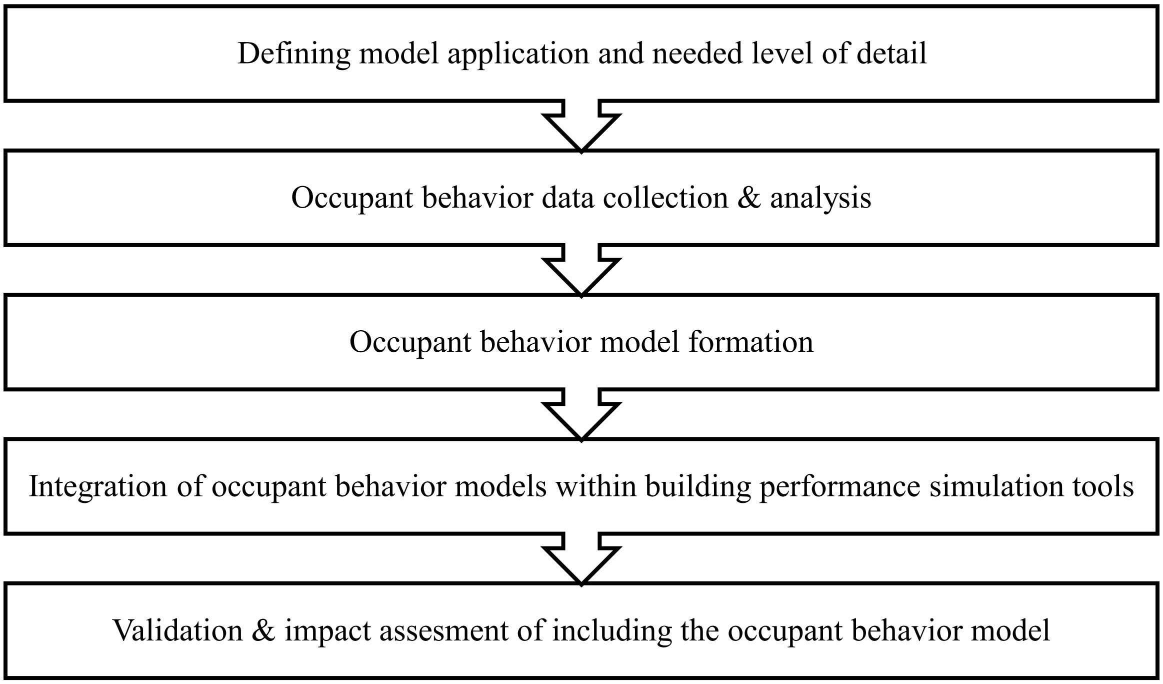


Figure 5. OB modeling, implementation, and validation sequence in BPSs.

According to Melfi et al. [103] there are three major factors that contribute to the resolution of an OB model, i.e., temporal, spatial, and occupancy. Temporal resolution is the level of accuracy used to model an event's timing and can vary from minutes and seconds to days and hours. Spatial resolution is correlated with accuracy on the physical scale, and it refers for example to the model's capability in predicting occupant numbers in a zone or building. How the model identifies individual occupants is referred to as occupant resolution, which ranges from models that are merely capable of determining if a space is occupied to models that are able to determine the precise behavior being performed by the occupant. The resolution for each of the three aspects must be well-defined before creating an OB model. It is also dependent on the objectives of the model and the issues it aims to solve. Figure 6 illustrates the resolution levels of OB models. Then, in an effort to comprehend the OB and learn more about the building's energy usage, data is gathered via sensing equipment and supplemented, if possible, with data obtained from the occupants themselves. Since OB models are meant to successfully explain the energy-related OB in buildings other than the ones from which the data was initially gathered, the model should be able to accurately describe the fundamental characteristics of OB while avoiding the inclusion of any irregularities particular to the data source. A model is considered suitable for widespread use (meaning that it can be used in multiple cases by modifying certain inputs) if it's robust, pragmatic, and has an acceptable number of inputs that can be easily defined [88].

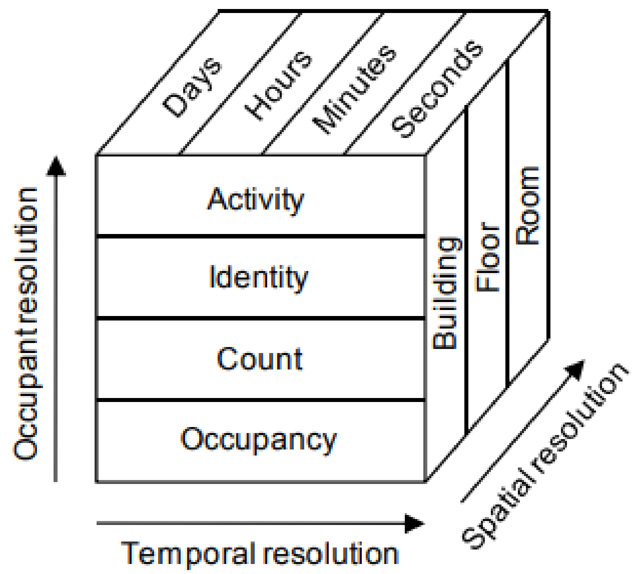


Figure 6. OB model resolution levels [103].

In general, occupants' energy-related behavior is caused by a variety of cues referred to as "drivers" [104], [105]. The relationship between drives and behavior, as well as behavioral patterns and what motivates them throughout time, can then be examined to develop a behavioral model. In this way, despite OB's complex, dynamic, and interdisciplinary nature, it may be represented by quantitative models. These models can then be integrated within BPS tools for a more reliable representation of OB. The uncertainty, associated with almost every step in OB modeling, also remains an important aspect to be considered. The inappropriate choice of OB data collection, modeling technique, and integration approach can all introduce more uncertainties in the process. OB models can be used in conjunction with statistical techniques such as uncertainty and sensitivity analysis to identify important factors affecting particular performance indicators (PIs), such as heating, cooling, and electricity demand. The variability of PIs owing to OB can be reduced by taking into account occupant-related uncertainty and defining PIs using probability distributions or predicted ranges [85], [95]. Accordingly, various OB patterns can be utilized to generate probability distributions for PIs, enabling more accurate predictions of building performance and its potential variation as demonstrated in [106], [107], [108] while minimizing OB-related uncertainty and maximizing the robustness of the simulation to OB.

Although OB research has been of great interest to researchers over the past decade and noticeable advancements have been made in several aspects of OB such as data collection methods, modeling approaches, and model implementation. However, the use of OB models in BPSs is still largely limited to researchers rather than practitioners and users. Besides, a performance gap between simulated and actual energy use still exists. When it comes to urban-scale energy analysis, two main limitations hindering the representation of OB can be identified.

Acquiring empirical statistics and large-scale occupant behavior data: More energy may be utilized in residential structures in urban regions as a result of the rapid modernization and accompanying lifestyle changes. For the development and implementation of urban policies, it is crucial to comprehend building energy performance beyond individual structures and at a wider city scale. In our review, however, only a few studies examined how OB relates to the large-scale building energy performance [109]. The challenge of acquiring reliable information on building occupancy at an urban scale is hindered by the expense and complexity of deploying numerous sensing devices throughout the whole urban area, while relying on constant occupancy profiles is not advised for urban scale applications. In order to enable further analysis in this area, more field surveys should be conducted due to the importance of empirical evidence and data to the study of

OB and the building energy performance [92], [110]. The widespread usage of urban sensing technologies, Internet of Things (IoT), and open city data can also be employed to better understand and describe OB. For the purpose of occupant tracking and locating, researchers have used datasets leveraging a variety of mobile signals, including cellular tower signals, GPS, Wi-Fi, and Bluetooth which do not require the installation of additional expensive sensors, cameras, and calibration equipment [111]. Further efforts should be directed towards data creation and sharing while standardizing data models schemas for representing the collected data in a consistent format. Ensuring interoperability and facilitating easy integration into various applications while developing practical applications is vital to showcase the real-world utility and benefits of the collected data, highlighting its value and potential in diverse domains and use cases [112].

Modeling diversity between occupants: In most cases, OB models represent all occupants in a similar manner. The data collected from each space, or each occupant is mixed together, thus, obscuring the differences among occupants. Due to the stochastic nature of OB, even in similar situations, different occupants may behave in different ways. The occupants' demographics such as age and gender also affect their actions. This diversity between occupants can only be represented in OB models if each occupant is considered individually [88]. Some research has also proposed an improved definition of “average” occupants based on categorizing occupants regarding activity level or typologies discovered by cluster analysis in efforts of increasing model accuracy [85].

For larger scale simulations, due to the diverse manner in which occupants of different buildings may behave, a significant impact on peak loads can be observed for instance due to varying thermostat setpoint preferences [113]. Happle et al. [114] generated several space-based diverse and non-diverse occupant presence models and evaluated their effects on district occupancy, energy demand, energy potentials, and centralized cooling supply system design. They highlighted the need to consider diverse and realistic occupancy profiles in UBEM to improve the accuracy of energy demand predictions and support informed decision-making for urban energy systems.

Wu et al. [115] proposed a Level of Detail (LoD) methodology to determine the proper level of occupant air-conditioning behavior modeling granularity considering applications related to district cooling. Based on their findings, different LoDs are recommended for different district cooling applications. Their analysis concluded that disregarding occupant diversity was only acceptable when assessing total district cooling demand. While for district cooling and thermal energy storage systems design, and for electricity-pricing strategies occupant diversity had to be considered. These efforts however need to be complemented with more research regarding various behaviors, building typologies, and applications. Furthermore, the selection of representative and sufficient sample sizes capable of representing diversity in OB should be analyzed [85].

2.5. Literature Review Summary

Surrogate ML models integrate elements of machine learning or statistical approaches with physics-based simulations to leverage the strengths of both approaches. Several studies have repeatedly shown that these hybrid models effectively strike a balance, harnessing the strengths inherent in both white-box and black-box models, and achieving better prediction results [9]. However, their widespread adoption faces limitations due to uncertainties in their development methods and the absence of widely used software for grey-box models [10]. Another issue, which limits the potential of UBEMs is related to simplistic or inaccurate assumptions regarding OB [11]. OB is usually disregarded in urban-scale energy analysis, with data unavailability being the main reason. This work presents a user-friendly decision support environment, leveraging a hybrid approach, combining representative building physics-based models and machine learning techniques to estimate the energy consumption of a mixed-use district. The tool is created by linking the trained ML models to an Excel sheet. This tool would enable the prediction of building cooling energy use intensity using easily definable inputs by the user. The results from these individual predictions are then aggregated to obtain the district-level cooling energy consumption profile. The diversity between buildings of the same type can be more accurately represented by including inputs related to occupant behavior.

CHAPTER 3

3. Methodology

3.1. Research Framework

The research methodology is detailed in this section and can be divided into four main steps as illustrated in Figure 7. The initial step involves the creation of representative building archetype models. The cooling load profiles of these archetypes are analyzed and validated using measured data from the case study district to ensure their representativeness and applicability. In the second step, the parametric simulation process used to generate the necessary datasets is described. An R script automates the creation of input parameter combinations, applies these parameters to the building archetype models, runs the simulations, and consolidates the input parameters with the required outputs in CSV files for further steps. In step three, the surrogate ML models are developed, detailing the training, testing, and optimization processes. Finally, in the last step, the proposed decision support environment is tested and deployed to predict the cooling loads for each building in the district in a user-friendly and computationally efficient manner, before aggregating the loads at the district scale. It should be noted that the district-level cooling load profile is developed by simply aggregating the cooling load profiles of individual buildings in the district. According to the ASHRAE DC Guidelines and the IDEA DC Guidelines, building-to-building interactions add unnecessary complexity at the initial design stage of DC plants. Therefore, simple aggregation is considered sufficient for initial design purposes.

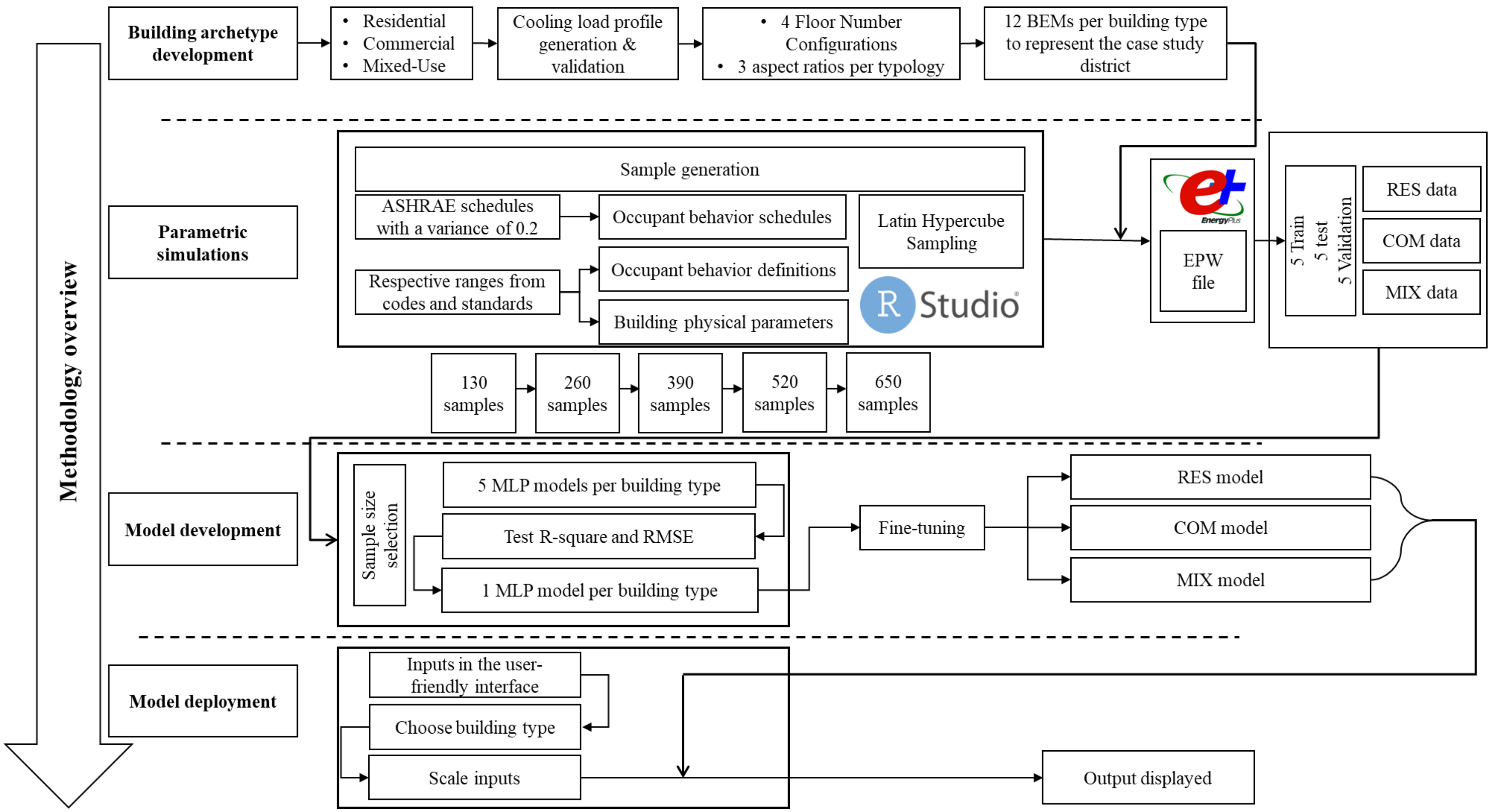


Figure 7. Methodology overview.

3.2. Case Study

The Marina district, which the case study of this research is built on, is considered the downtown core of Lusail City, Qatar. Lusail is recognized as Qatar's primary smart and sustainable city project, aligning with Qatar’s long-term aspirations for sustainable development. The Marina district contains over 100 plots of land, intended for the construction of high-rise towers for residential usage, office space, or mixed-use towers ranging from 12 to 50 stories. All buildings must adhere to the Global Sustainability Assessment System (GSAS) 2-star rating following local authority requirements. To meet the cooling requirements of the Marina district, a District Cooling Plant (DCP) with a capacity of 92,000 tons of refrigeration (TR) was installed. A three-dimensional aerial shot of the district is shown in Figure 8. The selection of the Marina district for this case study was based on several factors, including the region's extreme climatic conditions, its dependence on district cooling, the prevalence of high-rise buildings, compliance with the Lusail City GSAS 2 Star Rating Requirements, and the current sustainability trends.

Figure 8. 3D aerial view of the Marina District.

3.3. Building archetype development

A building archetype refers to a representation of a cluster of buildings within a specific area, sharing comparable attributes and specifications. Archetypal segmentation simplifies the task of representing or modeling numerous buildings, as archetypes can encapsulate the typical energy characteristics of the buildings they represent [116]. The development of building archetypes typically involves two primary steps: classification and characterization. In the classification step, the building stock is segmented into homogeneous groups based on one or more classifiers of energy behavior. The characterization step entails defining the thermophysical characteristics, HVAC system specifications, and inputs describing internal loads and schedules in accordance with applicable codes and standards as well as pertinent literature.

In most research, the main variables, or classifiers, used to categorize the building stock include the building geometry, building typology or usage, year of construction, HVAC system details, climatic zone of the building, and envelope thermophysical properties [117]. In the specific context of the Marina district, where all buildings adhere to the Lusail City GSAS 2 Star Rating Guidelines, were constructed during the same period, and are served by the district cooling plant, the primary factors considered for archetype classification were building typology and geometry. the Lusail City Development Guideline and the Marina District master plan were consulted to identify the predominant building usage typologies and geometrical variations. Three primary high-rise building typologies were identified: multi-unit residential buildings, commercial buildings comprising mainly office spaces, and mixed-use buildings incorporating residential, office, and retail spaces. Every floor in the residential archetypes is depicted as a single thermal zone, used for residential purposes. Comparably, for commercial archetypes, each floor is modeled as a thermal zone corresponding to office space. However, in the case of mixed-use archetypes, the allocation includes 10% of the space for retail, 20% for office usage, and the remaining 70% for residential purposes. Accordingly, three high-rise building archetypes were initially modeled, one representing each of the prevalent building typologies in the district. An aspect ratio of 1:1 was assumed for the three archetypes, as it was the most commonly identified aspect ratio in the Marina District development plan. As for the number of floors and gross floor area (GFA), mean values were calculated from all district buildings to define them, as presented in Table 1.

Table 1. Geometrical characteristics of the three initial building archetype models.

Archetype	**Commercial**	**Residential**	**Mixed-use**
Number of floors	28	19	31
Gross floor area (m^2)	28,672	15,979	29,791

For defining the internal loads, information from the Lusail City GSAS 2 Star Rating Guidelines was prioritized, being a region-specific guideline to which the district's buildings must adhere. When necessary, this guideline refers to ASHRAE or other applicable standards to obtain the required information. Table 2 illustrates the general building envelope requirements enforced for all buildings in the district, which were subsequently implemented in the archetypes according to the Lusail City GSAS 2 Star Rating Guidelines. A list of the internal load definitions used for the archetype development is presented in Table 2. The input parameters were initially gathered as ranges. Subsequently, their means were calculated and utilized in the development of the baseline building archetype models. Schedules for occupancy, lighting, plug loads, and infiltration were extracted from the "ASHRAE 90.1-2016 User's Manual" [118] since no region-specific schedule information was available. To ensure the accuracy and representativeness of the created archetypes and the validity of the inputs used for characterizing the archetypes, the cooling load profiles were simulated and analyzed for the three archetypes and validated against measured data and available literature.

Table 2. Building envelope requirements as per the Lusail City GSAS 2 Star Rating Guidelines

Parameter	**Value**	**Unit**
Wall U Value	≤0.3	W/m^2K
Roof U Value	≤0.25	W/m^2K
Window U Value	≤ 1.8	W/m^2K
Window/Wall Ratio	50	%
Window Solar Heat Gain Coefficient	≤ 0.25	NA
Cooling Setpoint	23	℃
Floor Height	4	m

Table 3. Internal load definitions for the baseline archetypes

Parameter	**Unit**	**Office**	**Residential**	**Retail**	**Reference**
People definition	People/m^2	0.0538	0.0283	0.1605	ASHRAE 62.1-2022, ASHRAE 90.1 -2016 User's manual
Lighting Power Density	W/m^2	9	6.5	9.038	GSAS 2 Star Rating Guidelines, ASHRAE 90.1-2019
Electric Equipment Power Density	W/m^2	8	6.67	3.228	ASHRAE 90.1-2016 User's manual
Infiltration Rate	m^3/m^2s	0.00057	0.00057	0.00057	PNNL-18898 guideline
Ventilation Rate	$m^3/s{\cdot}person$	0.0025	0.0025	0.0038	ASHRAE 62.1-2022

After analyzing and validating the results from the three initial building archetypes using available measurements and literature benchmarks, the Marina District development plan was used to identify the possible geometric variations of the buildings. In terms of geometric variations, four potential floor count values and three possible values for the buildings' aspect ratio for each building typology were identified. Thus, a total of 36 archetypes were developed, 12 archetypes for each building typology, to accurately represent the building stock. The geometrical characteristics of all the archetypes modeled for representing the district are summarized in Table 4.

Table 4. Geometrical variations for the archetypes representing the case study district.

	Residential	**Commercial**	**Mixed-Use**
Floor number	16, 19, 21, 26	25, 28, 31, 42	15, 23, 31, 40
Aspect ratio	1, 1.5, 2		
WWR	50 %		

The building energy models were developed using SketchUp and OpenStudio. A window-to-wall ratio (WWR), calculated as the percentage area obtained by dividing the building's glazed area by its wall area, of 50% was used for all models, following the Lusail city GSAS 2 Star Rating Guideline. Following the creation of the models' geometry, space types, and thermal zones were allocated to different areas, along with the associated details describing the construction, building activity, internal loads, and schedules. The simulations were then executed using EnergyPlus. For the weather file used in the simulations, a TMY weather file from the Doha International Airport weather station was chosen due to the close proximity between Doha and Lusail City and their similar coastal climates. For cooling load calculations, the Ideal Air Loads system in EnergyPlus is used.

3.4. Parametric simulations

In this step, the generated building archetype models are employed to generate the required datasets for training and evaluating the surrogate ML models. The variables considered for the parametric analysis can be broadly categorized into three groups: building thermophysical properties, occupant-related definitions describing building operation, and building operation schedules. For the first two categories, ranges are defined for each variable based on information gathered from relevant codes and standards, as outlined in Table 5. The selection of these variables in the parametric simulations was informed by a sensitivity analysis presented in a previous study conducted in the same case study district [81]. Due to the difficulty in determining the characteristics of the distribution of each variable between the set minimum and maximum values, a uniform distribution is used. Regarding the building operation schedules, using reference schedules proposed by standards like ASHRAE, which describe the aggregate temporal variations of internal gains, can result in repetitive internal load profiles across buildings of the same type, which may not accurately represent real-world diversity. To address this limitation and achieve a more realistic representation of occupant-related variables, probability density functions (PDFs) are defined for occupancy, lighting, and electric equipment usage schedules, allowing for the incorporation of stochastic schedules in the parametric simulations.

ASHRAE proposes standard daily schedules describing the rates of occupancy, lighting, and appliance usage for various building types. Given the absence of region-specific schedules, these schedules are used as a starting point for defining the PDFs and introducing the desired variability needed for the stochastic schedule sampling. To stochastically represent occupant-related factors for each building, randomized schedules were created based on these reference schedules. The proposed reference schedules consist of blocks of hourly periods covering the 24 hours of each day (e.g., [01:00 – 07:00], [10:00 – 16:00]). Accordingly, a probability distribution was generated for every hourly block based on the mean value suggested by the reference schedule for that block, along with a default coefficient of variance (CV). Following prior research on schedule diversification procedures, a CV of 0.2 was employed to generate the needed PDFs [44], [45]. Typical residential building weekday schedules, as well as the stochastic schedules employed, are illustrated in Figure 9.

Table 5. Variable input ranges for the parametric simulations.

Parameter	Unit	Residential	Office	Retail	Reference
Wall Thermal Resistance	m^2KW^{-1}	1.75 – 3.33	1.75 – 3.33	1.75 – 3.33	Qatar Construction Standard / ASHRAE 90.1 2019
Roof Thermal Resistance	m^2KW^{-1}	2.27 – 4.5	2.27 – 4.5	2.27 – 4.5	Qatar Construction Standard / ASHRAE 90.1 2019
Window U Value	$Wm^{-2}K^{-1}$	1.8 - 1.9	1.8 - 1.9	1.8 - 1.9	GSAS 2 Star Rating Guideline / Qatar Construction Standard
Solar Heat Gain Coefficient	NA	0.25 – 0.275	0.25 – 0.275	0.25 – 0.275	Qatar Construction Specifications / ASHRAE 90.1 2019
Occupant Density	m^2/person	38 - 90	1.67 - 50	2.5 – 12.5	ASHRAE 62.1 2019
Lighting Power Density	Wm^{-2}	1 – 6.5	8.5 – 9.6	3 - 9.5	GSAS 2 Star Rating Guidelines / ASHRAE 90.1-2019
Equipment Power Density	Wm^{-2}	2 – 8	2.5 – 21.53	2.5 – 21.53	ASHRAE 90.1-2016 User's manual
OA Ventilation Rate	$ls^{-1}m^{-2}$	0.3 – 0.5	0.3 – 0.6	0.3 – 0.9	ASHRAE 62.1 2019
Infiltration Rate	ACH	0.1 – 0.2	0.1 – 0.2	0.1 – 0.2	ASHRAE Fundamentals
Cooling Setpoint Temperature	℃	23 - 26	23 - 26	23 – 26	GSAS 2 Star Rating Guidelines

The Latin Hypercube Sampling (LHS) technique [119] is utilized to create randomized sets of input parameter combinations from the specified ranges for the various parameters. LHS is a statistical approach that stands out as a favorable choice for sampling procedures, especially when dealing with computationally demanding models [120]. This approach allows for a reduction in sample size while producing more reliable results compared to random sampling. An R script utilizing the EplusR library [121] is then utilized to automate several steps in the procedure. First, an input parameter combination is created by sampling a value for each of the input parameters using the predefined PDFs. Secondly, the input parameter combination is passed on to the building archetype seed models as measures, meaning that new building energy models are created using the generated input parameter combination. For every input parameter combination, this process is repeated for all 12 archetypes of the respective building type. The created EnergyPlus models are then simulated to generate the monthly cooling energy consumption. The input parameter combinations, building type, floor number, and building aspect ratio along with the corresponding monthly cooling energy consumption for each one are then consolidated in a CSV file to be used in further steps. Additionally, the average monthly dry bulb temperature and average monthly relative humidity are extracted from the TMY weather file as shown in Table 6 and included in the CSV file as well. Previous research has shown that these two weather indicators sufficiently capture climate conditions in hot and humid regions [122], making them essential for developing surrogate ML models. The number of input parameter combinations determines how much this process would be repeated and the associated computational needs.

Table 6. Weather indicators for describing the case study climactic conditions.

Parameter		Reference
Monthly outdoor air dry-bulb temperature (℃)	19.06, 19.26, 22.21, 27.30, 32.29, 34.77, 36.07, 35.78, 33.90, 30.67, 25.80, 21.30	TMY weather data for Doha
Monthly outdoor air relative humidity (%)	63.20, 59.15, 62.04, 50.07, 38.48, 41.77, 36.07, 56.85, 52.27, 57.40, 52.56, 61.15	TMY weather data for Doha

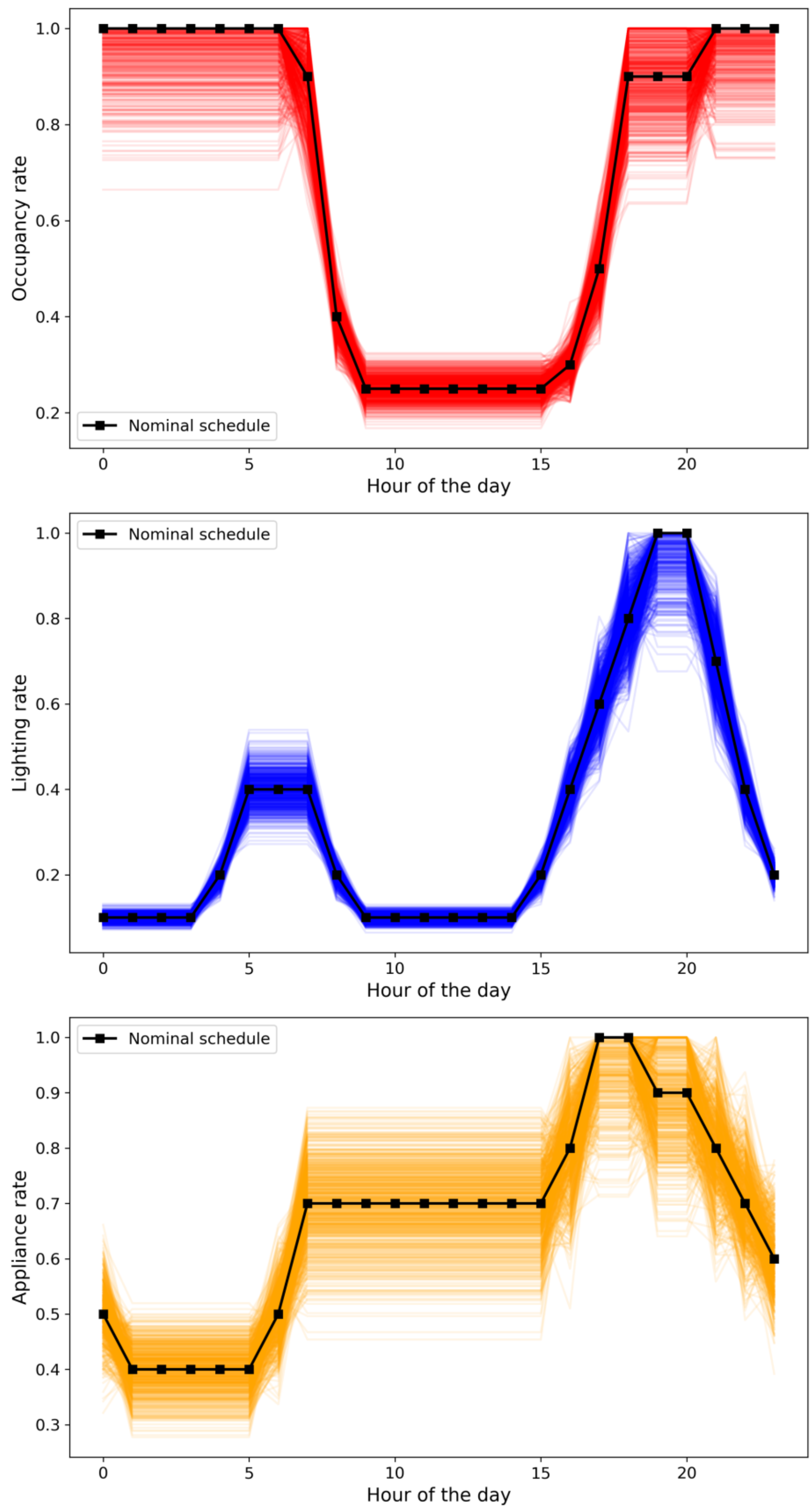


Figure 9. Residential building weekday reference and stochastic schedules.

Determining the appropriate number of samples for surrogate model creation remains an open research question [123]. The surrogate ML models necessitate substantial volumes of data for training, yet beyond a certain threshold, the inclusion of more data points might not contribute significantly and could, in some instances, be detrimental if it causes model overfitting. A general rule of thumb suggests that at least 10 LHS samples would be required for each sampled variable in the surrogate ML model [124], [125]. For this study, given that 13 variables are considered in the parametric simulations, the number of samples (input parameter combinations) should not be less than 130. To examine how the dataset size might impact the performance of the surrogate ML models, and accordingly determine the proper dataset size for each model, the augmented LHS function [126] was used to create 5 separate data sets with sizes of 130, 260, 390, 520, and 650 samples which are used for training the ML models. The augmented LHS function attempts to add points to the design in an optimal manner, by adding points to an existing LHS while preserving the Latin properties of the design [126]. Accordingly, a fair comparison can be made based on the sample size and not the sample properties. Additionally, to ensure the representativeness of the data sets, separate testing and validation sets are created, each corresponding to 10 % of the training set size, following the same approach. This process is performed for each of the 3 building typologies.

3.5. Model development

The development of the ML models involved various steps. First, using the input-output data sets created from the parametric simulations, the data was pre-processed and normalized. The ML models are constructed using a feedforward ANN algorithm, specifically the multilayer perceptron (MLP). After completing the data pre-processing procedures, five models are trained for each building type, utilizing the different dataset sizes. The performance of each of these models is then evaluated using both the test and validation sets, ensuring a robust analysis. Evaluation metrics such as the coefficient of determination (R^2) and the root mean square error (RMSE) are utilized. Based on these results, the appropriate dataset size is selected to establish one MLP model per building typology. Then, a grid search is conducted to identify the optimal hyperparameters for each model. Following the grid search, the models undergo final training, employing an early stop mechanism to prevent the finalized models from overfitting the training data, and integrating the selected hyperparameters. Lastly, the finalized models are tested and evaluated using various performance metrics.

3.5.1. Data pre-processing

The data pre-processing stage involves two main steps. First, to enable comparative analysis, the cooling energy consumption outputs are normalized by the floor area of the respective building archetype models. This normalization process converts the outputs into energy intensities, specifically monthly cooling EUIs. These energy intensities serve as the target variables during the development of the ML models. Additionally, considering the target variable as the monthly cooling EUI, the stochastic schedules for each simulation are factored in via a weighted average to represent the average monthly occupancy, lighting usage, and appliance usage rates.

Secondly, data normalization is particularly important when dealing with input data that exhibits significant variations. Data normalization is a crucial preprocessing step employed to scale data proportionally within a specific interval. This is essential to mitigate the influence of magnitude and units of the predicting variables, which could otherwise disrupt the model fitting process. A zero-mean normalization (standardization) process was used for normalizing the predicting variables. This process normalizes the data by adjusting it to have a mean (μ) of zero and a standard deviation (σ) of one as shown in Eq. 1.

$$x' = \frac{x - \mu}{\sigma} \qquad (1)$$

Where x' is the new normalized value, x is the original value, μ is the mean value, and σ is the standard deviation.

3.5.2. Artificial neural network (ANN)

An ANN is a computational model that mimics the structure and function of a biological nerve cell. It consists of interconnected nodes, called neurons, arranged in layers. Each neuron receives input signals from the neurons in the previous layer, processes this information using a transfer function, and produces an output signal that is transmitted to the neurons in the next layer through weighted connections, to create output data that are transmitted to the neurons that follow. The weights on these connections determine the strength of the influence of one neuron on another. During the training process, the ANN learns from input-output pairs provided by the training data. This iterative procedure adjusts the weights of the connections between neurons to minimize a predefined parameter, such as the sum of squared errors (SSE) or the root mean squared error (RMSE). By repeatedly adjusting these weights based on the errors between predicted and actual outputs, the network improves its ability to accurately map inputs to corresponding outputs. The training process concludes when a certain condition is met, such as reaching a predetermined maximum number of iterations, known as epochs. The model structure for an ANN is shown in Figure 10.

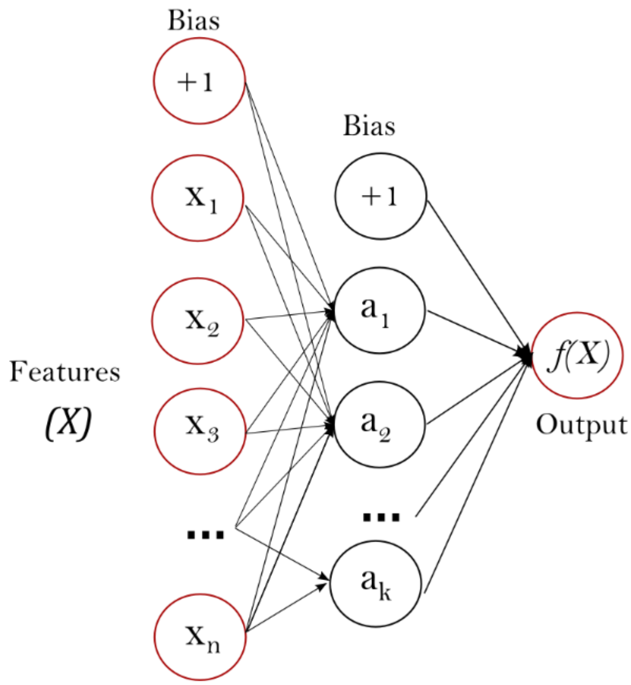


Figure 10. ANN structure with one hidden layer [127].

The most straightforward and commonly used architecture for ANNs is the feed-forward multi-layer perceptron (MLP). It comprises an input layer that receives external data (independent variables), an output layer that delivers the prediction results (dependent variables), and one or more intermediate hidden layers connecting the input and output layers. In this study, a feed-forward MLP model is employed, featuring three layers, including one hidden layer. To determine the appropriate sample size for each of the final models, 5 MLP models are trained using the different training set sizes, for each building type. For these models, a rule of thumb is used for determining the number of neurons for the hidden layer as shown in Eq. 2 [128].

$$n_H = \frac{2}{3}(n_I + n_O) \qquad (2)$$

Where n_H is the number of neurons in the hidden layer, n_I is the number of neurons in the input layer, and n_O is the number of neurons in the output layer. Accordingly, since the input layer contains 17 neurons, and the output layer 1, 12 neurons are initially used for training the models. The models are trained until either the loss or score stabilizes or when the maximum number of epochs, set at 300, is attained.

3.5.3. Performance evaluation

The performance of regression-based ML models can be evaluated using various statistical metrics, which gauge precision, accuracy, and generalizability. For this study, the prediction accuracy of the models was assessed mainly via two metrics, the RMSE and the coefficient of determination (R^2), as shown in Eq. 3 and Eq. 4 respectively. The RMSE can evaluate both the bias and the variance of the predicted values compared to the measured output, while R^2 indicates the goodness of fit of the model. A higher R^2 value indicates that the regression model fits the data better, meaning that its predicted values are closer to the actual observed values.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \qquad (3)$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \mu)^2} \qquad (4)$$

Where y_i and $\hat{y}_i$ represent the actual and predicted values respectively, n is the number of samples and μ refers to the mean of all values.

3.5.4. Sample size selection

Since five different models for each building typology were trained using various augmented dataset sizes, evaluating the impact of dataset size on model accuracy is crucial for determining the appropriate dataset size. This evaluation helps inform decisions about balancing computational efficiency and model performance. Generally, The ML models require sufficiently large datasets

for training. However, beyond a certain threshold, adding more data points may not improve model performance and could even be detrimental, potentially leading to overfitting.
Each of the trained models was evaluated using both test and validation sets, based on the R^2 and RMSE metrics. Using both sets allowed for a more robust performance evaluation, and the results were averaged to provide a comprehensive assessment of model accuracy and generalizability. A trade-off was established to select the appropriate dataset size to reach one MLP model per building. For this study, it was decided that no additional samples would be needed for model training if performance decreased with increasing sample size, indicating overfitting, or if the improvement in RMSE was less than 3%. A performance increase of less than 3% was deemed insufficient to justify the computational burden of generating more data points for this application. Using this criterion, an appropriate sample size is selected for each building typology, to reach one MLP per building type.

3.5.5. Hyperparameter tuning

Hyperparameters are adjustable parameters that significantly influence the performance of ML models. Hyperparameter optimization or tuning involves finding a set of model hyperparameters that ensures optimal performance of the model, by enabling the model to better reflect the relationship between inputs and outputs. For this study, a full grid search technique is performed for tuning the hyperparameters of the models using the separate validation sets. The use of a separate validation set instead of the more traditional *k*-fold cross-validation process, decreases the computational burden, thus allowing for an exhausting search of all hyperparameter options. Also, given that the training sample is generated using LHS, the representativeness of the data set would have been compromised if divided into *k*-folds [123]. The hyperparameters used in the grid search process are shown in Table 7.

Table 7. hyperparameter values for grid search.

Hyperparameter	**Tested values**
'hidden_layer_sizes'	(4,), (16,), (35,), (245,)
'activation'	'identity', 'relu', 'tanh', 'logistic'
'solver'	'adam', 'sgd', 'rmsprop'
'alpha'	0.0001, 0.001, 0.01
'learning_rate'	'constant', 'invscaling', 'adaptive'
'batch_size'	64, 128, 256

3.5.6. Final model training

After the optimal hyperparameters are configured for each of the three models, a final version is trained using the selected sample size for each building typology and the optimal hyperparameters. An early stop mechanism [129] is employed in the final training process to avoid overfitting the training data. This procedure consists of several steps. First, an additional validation set is created for each of the three models and a maximum number of iterations, 300 in this case, is defined for the early stop procedure. Secondly, the training process commences, and the model's performance is evaluated over each iteration using the validation set after updating the weights and biases. If the model performance deteriorates over a predefined number of epochs, 5 in this case, the model training ends. Finally, the weights and biases for the iteration corresponding to the best performance on the validation set are adopted for the model. An illustration of the early-stop mechanism as presented in [123] is shown in Figure 11. Following the training of the final models, their performance is evaluated once more using the corresponding test sets based on the R^2 and RMSE.

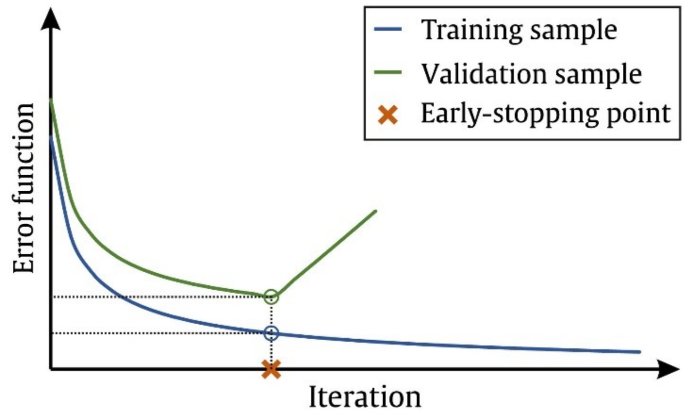


Figure 11. Early-stop mechanism [123].

3.6. Model deployment

After developing the ANN models, a user-friendly decision support environment (interface) was created to enable straightforward and efficient prediction of the cooling EUIs of different buildings within the district, without requiring expertise in BPS tools or programming. This interface was developed in several steps. First, the scalers used for normalizing the data in the pre-processing phase, along with the final trained ML models are saved using the 'Joblib' [130] python package. Secondly, a Python script is written to load the saved scalers and pre-trained models and eventually make the predictions given a set of inputs. This Python script is linked to an Excel sheet using the 'xlwings' [131] library. 'xlwings' is a Python package that makes it easy to call Python from Excel and vice versa. It allows users to interact with Excel spreadsheets using Python, enabling them to automate data processing, create reports, and perform complex calculations. In the Excel sheet, users choose the building type through a drop-down menu (residential, commercial, or mixed-use) and input a value for each of the 17 input parameters depicted in Figure 12. This information is then passed to the Python script, which scales the data and calls the appropriate ML model based on the defined building type, makes the prediction, and sends it back as an output in the Excel sheet.

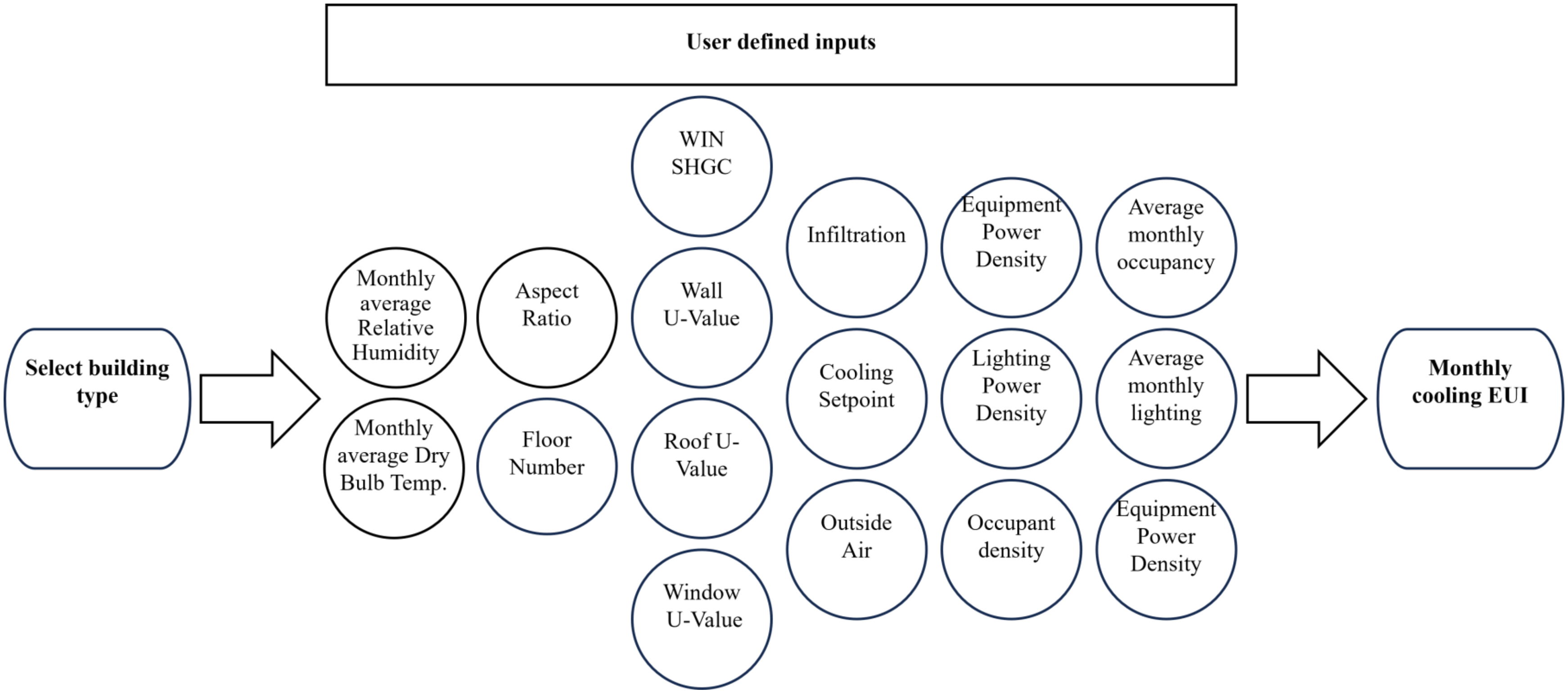


Figure 12. Inputs and output of the user-friendly decision support tool.

3.7. Summary

In this research, a surrogate ML UBEM is proposed. Initially, three physics-based archetypes, representing the main building typologies in the case study district, are developed. The cooling load profiles are simulated, analyzed, and validated for these archetypes before aggregating the results to the district scale. This step ensures the accuracy of the inputs used for characterizing the archetypes and demonstrates a common method for urban-scale building energy analysis. These archetypes serve as the foundation for creating surrogate ML models. The archetypes are expanded to consider different geometrical variations within the district, resulting in a total of 36 archetypes (12 per building typology). Data for training and testing the ML models is generated by automating parametric simulations using these archetypes. The reliance on artificial data for developing the ML models is due to the lack of measured data, which is especially common in the early design phase of urban regions. The created datasets are then used to train, test, and optimize the performance of three ML models, one for each building typology, capable of accurately predicting the monthly cooling EUI for any building in the district. Following the development of the ML models, a decision support environment (interface) is created by linking the trained ML models to a user-friendly Excel sheet. This interface allows users to input the characteristics of each building in the district to generate the cooling EUI profile for each building. These profiles are then used to create the cooling energy consumption profile for each building based on its gross floor area. Finally, by aggregating these building-level profiles, a district-level cooling energy consumption profile is generated.

CHAPTER 4

4. Results and discussion

4.1. Urban scale cooling load prediction of high-rise buildings in a hot and arid climate – An archetype-based approach

This section addresses the performance evaluation of the three initial archetypes modeled to represent the case study district. As discussed in the methodology section, the archetype geometries were developed using the SketchUp plugin in OpenStudio. They were then characterized by defining all the necessary input parameters in OpenStudio before running the simulations via EnergyPlus software. For an efficient DCP design and operation, detailed and accurate building cooling load profiles are required to be estimated. Besides, heat gain components of buildings should be investigated to analyze the saving potentials of building cooling. In the literature, despite the availability of studies that analyzed building energy consumption under different climatic zones, a new contribution is still required to estimate building cooling load profiles at the district level. Further, heat gain components of each building archetype should be analyzed to provide insights into the cooling requirements of buildings. To that end, this study employed an archetype-based approach to predict and analyze building cooling load and heat gain components at a district level in the hot and arid climate, using a case study of the Marina district of Lusail City, Qatar. Three high-rise building archetypes are developed to represent the Commercial, Residential, and Mixed-Use building stock of the district. The developed archetypes for this study contribute to the establishment of a representative building archetype library for Qatar and other regions under the same climatic conditions. The cooling load component and cooling energy demand profiles for the archetypes are then simulated and analyzed at multiple temporal resolutions. The obtained results are also validated using measured cooling energy data supplied from the district cooling plant, as well as information from available literature before aggregating the results on the district scale. The developed cooling load profiles are imperative for the design and operation of the DCP and associated facilities. Further, by analyzing the contribution of each cooling load component at different time instances, energy conservation measures could be assessed realistically.

4.1.1. Annual cooling energy demand profiles

The daily variation of the cooling energy demand for the three building archetypes is plotted over a one-year period, along with the variation in daily average outdoor dry-bulb and wet-bulb temperatures in Figure 13. The daily cooling demand for the archetypes ranged from 487 kWh to 30273 kWh, 1050 kWh to 18720 kWh, and 2198 kWh to 32070 kWh for the commercial, residential, and mixed-use archetypes, respectively. A similar trend in the cooling energy demand variation is noticed for the three building archetypes despite differences in their usage, input parameters, and cooling demand values. Besides that, a similar trend observed between the temperatures and building cooling demand shows that the building cooling demand heavily relies on outdoor climatic conditions. The weather in Qatar can be classified as extremely hot in summer from May to October with more moderate weather in the winter from December to February whereas March, April, and November are transition months. Therefore, high cooling energy demands are observed in the hot summer season whereas the transition months and winter period show mild and low cooling demands, respectively. The correlation between the cooling demand and outdoor climate was demonstrated in a previous study by Gastli et al. [132] where it was reported that there is a linear correlation between the daily maximum electric demand and daily maximum temperature in Qatar at temperatures above 22 °C and attributed the increased electric demand to increased air conditioning load.

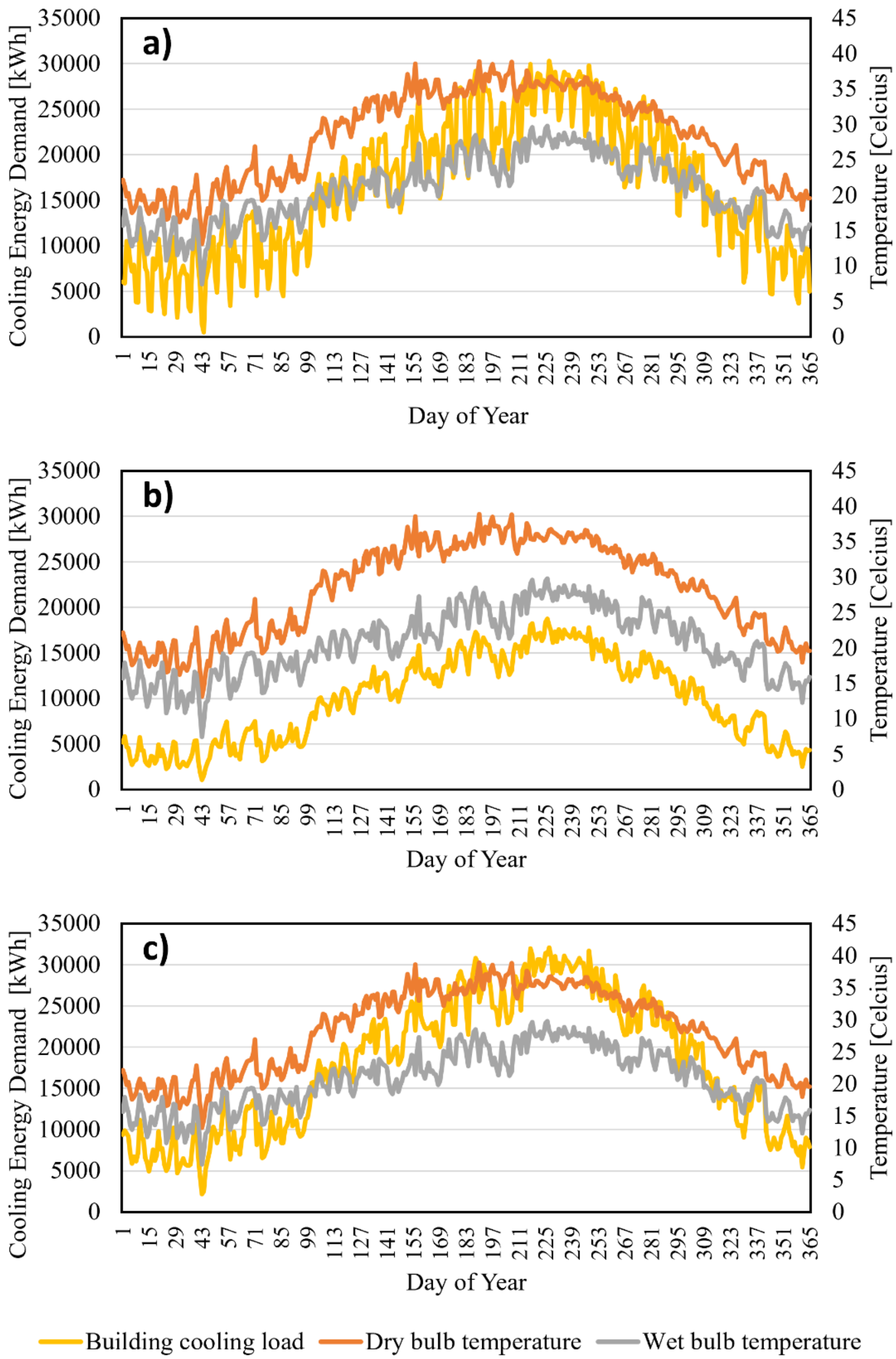


Figure 13. Annual cooling energy demand profiles for a) commercial, b) residential, and c) mixed-use archetypes.

4.1.2. Monthly cooling load profiles

The monthly cooling load breakdown for each archetype is presented in Figure 14. The results indicate that the heat gained from ventilation, infiltration, and convection represents the largest weather-dependent contributors to the cooling load. On the other hand, the use of deterministic occupancy, lighting, and electric equipment usage schedules results in a fixed heat gain from these components throughout the year which is independent of the outdoor climate. The use of accurate representative schedules would lead to high prediction accuracy. In our case, however, the lack of occupant-centric data hindered the use of region-specific schedules or occupant behavior models. Besides, the aggregation of the cooling load on the urban scale would account for some of the occupant behavior features which allows the safe assumption of the valid usage of these schedules at this design stage. For all three archetypes, the highest simulated cooling load is in August since the outdoor temperature reaches its maximum values during this month while the lowest cooling load is in January. For August, the largest cooling load component for the three archetypes was ventilation since introducing adequate outdoor air rates as per ASHRAE standard 62.1 requires a significant amount of cooling to lower the ventilation air temperature from the high outdoor air temperature to the indoor setpoint temperature. The heat gain from convection and infiltration was the second and third highest contributors to the cooling load, respectively, which is also due to the difference between the outdoor air temperature and indoor temperature setpoint. The electric equipment heat gain was the fourth largest contributor followed by the heat gain from solar radiation whereas the people and light were the lowest contributors to the cooling load. For January, the largest cooling load component was found to be the window solar radiation followed by the equipment heat gain and convection heat gain. The heat gains from occupants and lights come next in order where their values are fairly close while ventilation and infiltration have almost no contribution to the cooling load. This can be attributed to the moderate winter temperatures in the region which is close to the indoor temperature setpoint.

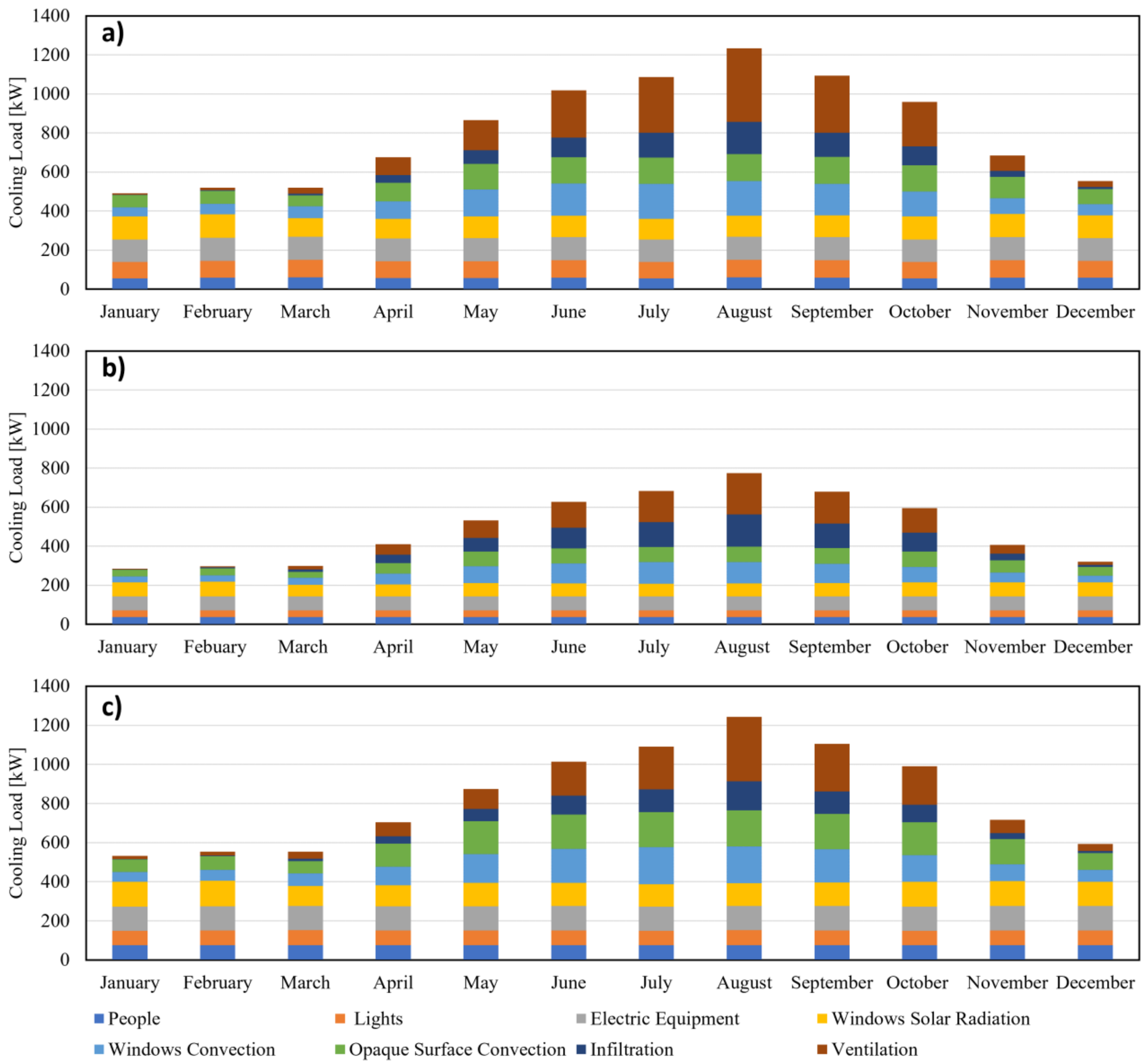


Figure 14. Monthly cooling load component breakdown for a) commercial, b) residential, and c) mixed-use building archetypes.

The cooling load is found to vary by a factor of approximately 2.5 from January to August among the archetypes. This large variation not only reduces the seasonal efficiency of DCPs but also requires a larger cooling capacity associated with an increased initial and operational cost.Several studies have demonstrated that Energy Conservation Measures (ECMs) can reduce these weather-dependent cooling loads and increase the building's resilience to outdoor climatic conditions. For instance, Krarti et al. [133] conducted an analysis on the effect of ECMs such as decreasing the infiltration rate by building a tighter building envelope, increasing the cooling setpoint which would reduce the ventilation load, and increasing the thermal insulation of the opaque building material which would result in less convective heat gains along with other ECMs to optimize the building energy consumption. They reported a possible energy saving of up to 47% in the case of Qatar by following the recommended ECMs. In another study by Andric et al. [134], increasing the building envelope insulation or the cooling setpoint could lead to energy savings of up to 30% annually. Further, Ahmed et al. [135] and Ortiz et al. [136] demonstrated in their case studies in New York city that changes in climatic variables have a considerable impact on building energy demand, with HVAC requirements being the main driver behind this increase. It was also mentioned that in some situations, such as during heatwaves, which are very likely to occur in the hot and arid climate of our case study region, the electric grid may struggle to satisfy peak demands, which are mostly driven by building HVAC needs. Hence, buildings should be designed and constructed to withstand the potential effects of climate change and be resilient to the extreme outdoor climate. The results of this study, along with the findings in the above-summarized literature, suggest the possibility of increasing building resilience to outdoor climate and illustrate the potential energy savings. Accordingly, the design, construction, and operation of buildings in hot and arid regions should target increasing building resilience to the outdoor climate and lessening the seasonal variation in cooling loads.

4.1.3. Hourly cooling load profiles

Figure 15 demonstrates the hourly breakdown of the cooling load components for the summer design day "July 21st" which is defined by the weather file used for the simulation.

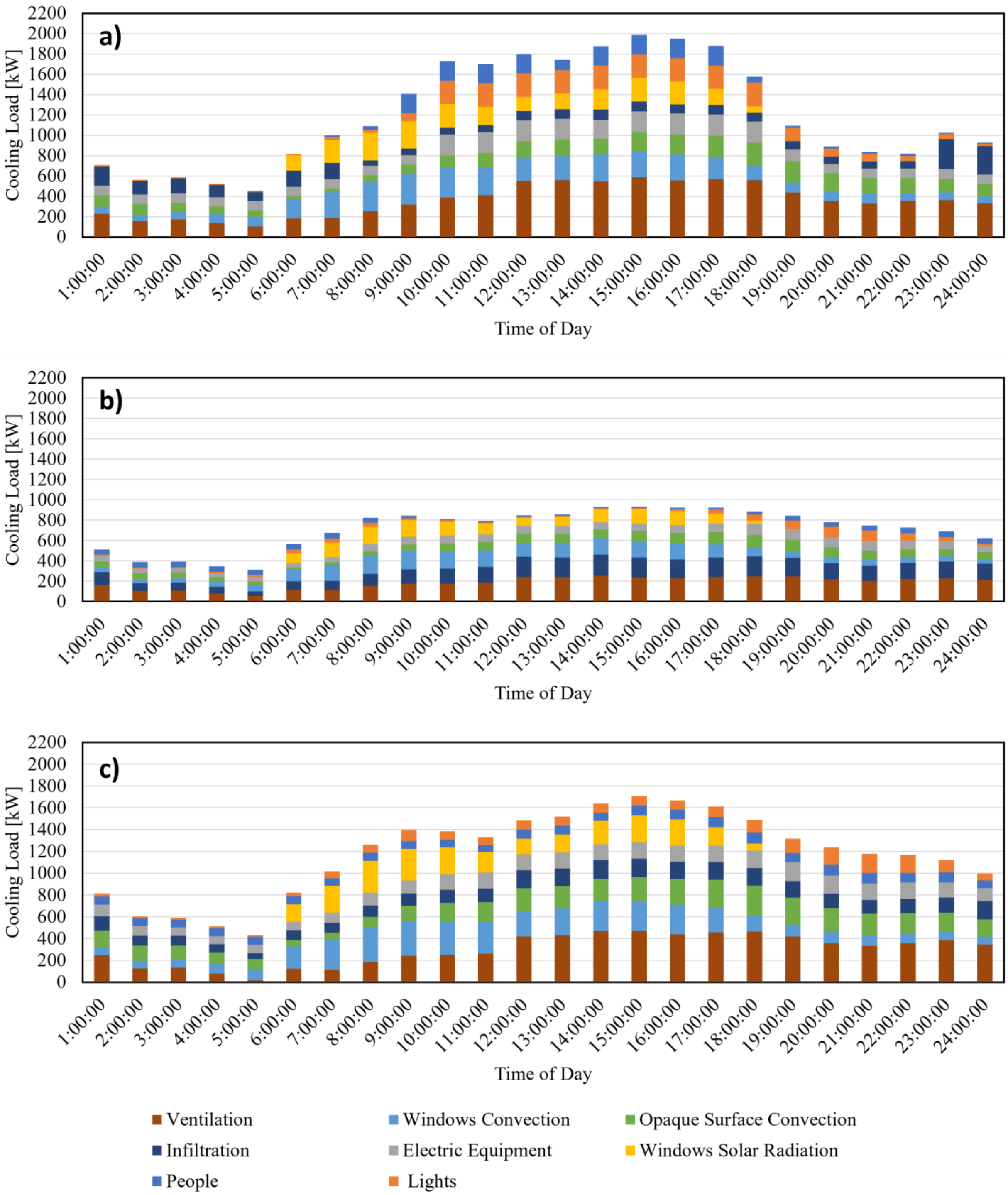


Figure 15. Summer design day cooling load component breakdown for a) commercial, b) residential, and c) mixed-use building archetypes.

Besides their role in DCP design and operation, these detailed profiles allow for investigating the potential for energy saving and peak load reduction of DCPs. For example, based on the results, solutions for minimizing ventilation rates at peak cooling loads could yield considerable energy savings. In that regard, Solgi et al. [137] proposed that considerable cooling energy savings can be achieved using night purge ventilation with phase change materials for the hot and arid climate of Iran. In the summer design day cooling load investigations, the ventilation, convection, and infiltration components of the cooling loads are observed to be highly affected by the outdoor temperature, resulting in a peak cooling load of 1986 kW, 932 kW, and 1704 kW at 3 pm and a minimum cooling load of 455 kW, 314 kW, and 432 kW at 5 am for the commercial, residential, and mixed-use building archetypes, respectively. It is noticed that the residential building displays the least variation in cooling load throughout the day, whereas the commercial building displays the highest variation. On the other hand, the mixed-use building displays relatively moderate cooling load variation. The observed cooling load variation can partially be attributed to the lower occupancy rate and the low variability in operation schedules of residential buildings. Residential buildings are modeled to almost always be occupied by a certain percentage; however, their low occupancy rates not only contribute less to the "people" cooling load as opposed to commercial buildings but also have less of an impact on the "ventilation" load since a ventilation rate per person is introduced in the building following the occupancy schedule. The commercial archetype, on the other hand, is characterized by high occupancy rates, which along with the high temperatures, contribute to the increase of the ventilation load when occupied. The mixed-use building falls between both in that regard since it's composed of retail, commercial, and residential spaces. Another noticeable observation is the high impact of infiltration in the cooling load of the residential building even though the residential building has the smallest GFA, the lowest cooling load, and the fact that all three archetypes have been modeled with the same value for the infiltration rates as a function of GFA. This is associated with the used infiltration schedules in the BEMs. For commercial and retail spaces, the higher ventilation rates, when occupied, could pressurize the building slightly, which would decrease the infiltration rates and, thus, the cooling needs associated with it. In residential spaces, however, the used schedules propose a constant infiltration rate, which leads to an increased cooling load associated with the infiltration. Solar radiation contributes to the cooling load from sunrise to sunset in the three archetypes, while the contribution of electric equipment, people, and lights varies depending on the building usage and operation schedule, causing, for example, the commercial building to have a higher contribution to the cooling load from people and lights when occupied compared to the other two building types.

In Figure 16, the hourly breakdown of the cooling load components is presented for the winter design day "February 21st". As in the case of the summer design day, the peak cooling load is reached at 3 pm for the three archetypes with values reaching 1176 kW, 526 kW, and 1081 kW, while the minimum cooling load was at 5 am with values of 105 kW, 110 kW, and 169 kW for the commercial, residential, and mixed-use archetypes, respectively. According to the results, solar radiation and electric equipment are the largest contributors to the cooling load on a winter design day, each contributing to approximately a quarter of the cooling load throughout the day. The window convection, along with the solar radiation, contributes to the cooling load during the daytime only from sunrise to sunset, while the opaque building material convection continues into the night due to the thermal mass of the building envelope. The cooling load contribution of the occupants and light follow the building operation schedule with higher people and light load in the commercial building when occupied, whereas a lower and more evenly distributed load is observed in the case of the residential and mixed-use buildings. On the other hand, ventilation and infiltration have minimal contribution to the cooling load due to the moderate outdoor climate in winter, which allows for natural cooling of the building and sizable energy savings throughout the winter season.

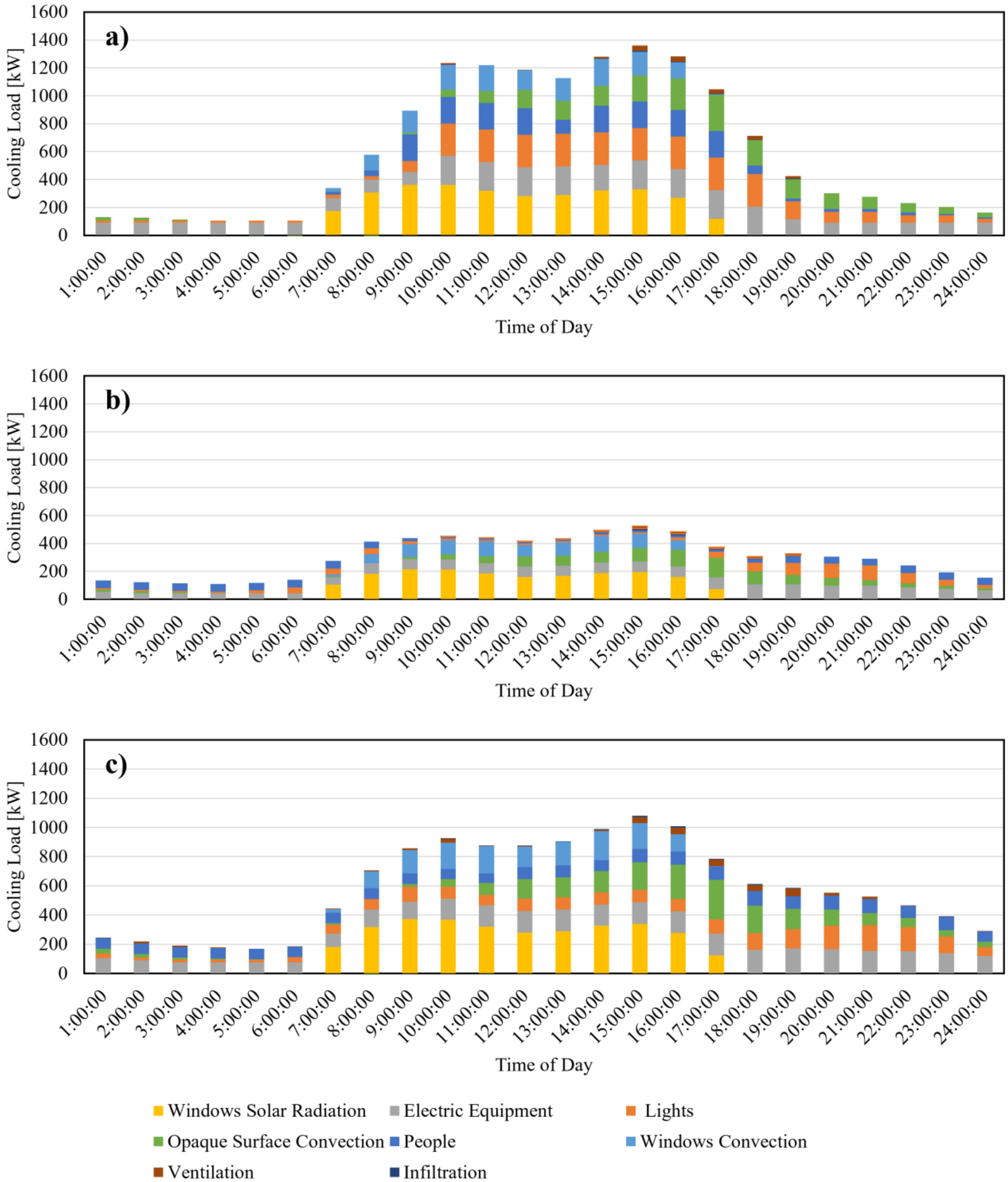


Figure 16. Winter design day cooling load component breakdown for a) commercial, b) residential, and c) mixed-use building archetypes.

4.1.4. Validation of the simulation results

For validating the results obtained from the simulation of the building archetype models, measured data regarding the delivered cooling energy from the DCP to a residential and mixed-use building was obtained and compared to the simulated results, while the commercial archetype was validated against available benchmarks in the literature. For the commercial archetype, the annual Energy Use Intensity (EUI) was used as a metric for evaluating the archetype performance since no region-specific cooling EUI was reported in the literature.

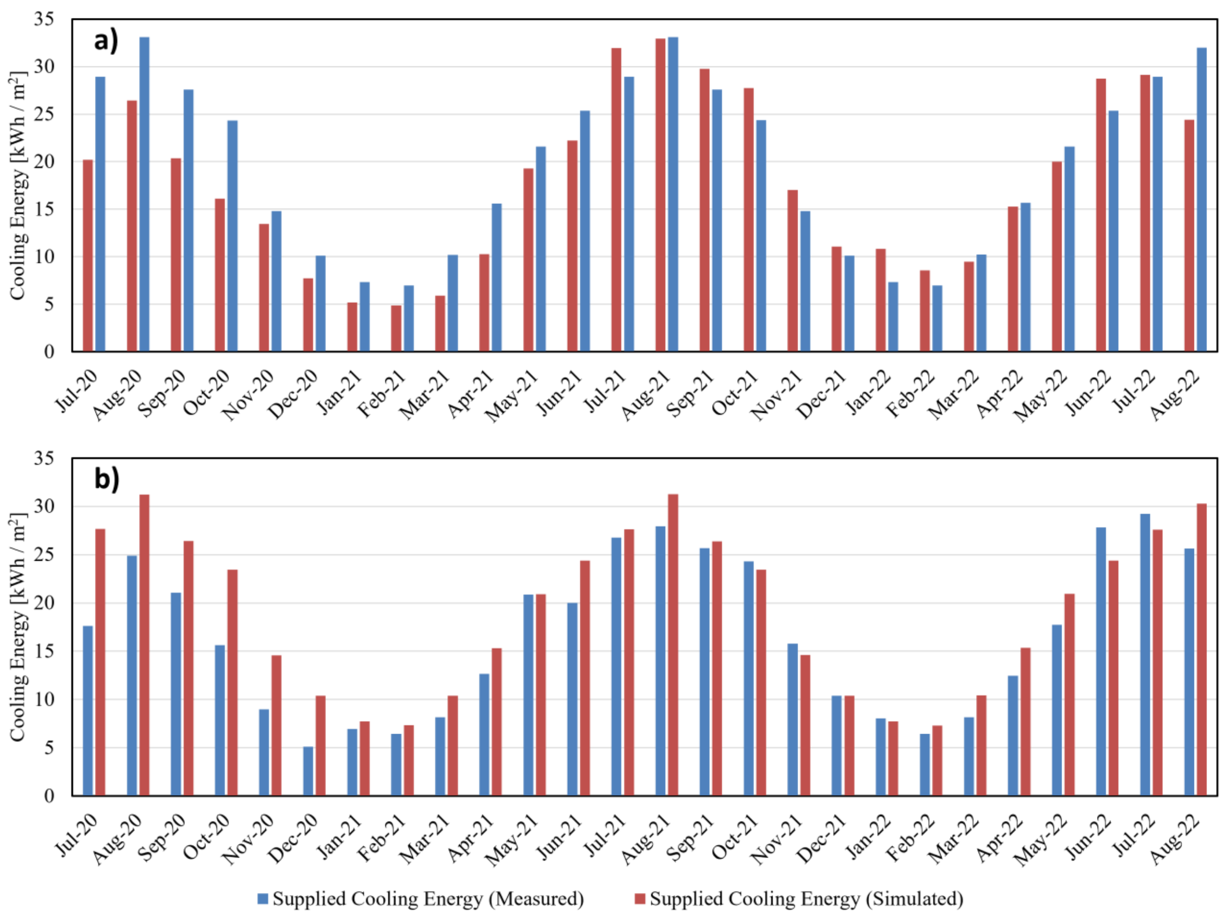


Figure 17. comparison of measured and simulated cooling energy demand for a) residential, and b) mixed-use building archetypes.

The annual EUI for the commercial archetype was found to be 272 $kWh/m^2/year$, which is similar to the reported EUI values in the region. For instance, Azar et al. [138] reported an average EUI value of 284 kWh/m2/year for commercial buildings in Kuwait. Another study by Alkaabi et al. [139] reported a mean of 280.4 kWh/m2/year for medium to high-rise office buildings in the UAE. The comparison between the measured and simulated results for the residential and mixed-use buildings is presented in Figure 17. A relatively large variation between the measured and simulated results is noticed from July 2020 to April 2021 since this period is the commissioning phase of the DCP and buildings are not in steady operation with a defined usage schedule. After that period, the simulated results match well with the measurements, which further supports that the archetype models are representative of the building stocks in the district.

4.1.5. District-level cooling energy demand profile

Table 8 presents the total floor area of each building stock in the district, as well as the cooling EUIs for the three archetypes. By multiplying the EUI for each archetype with the corresponding total floor area of building stocks, the annual cooling energy demand for each building stock was obtained. The summation of the annual cooling energy demand of the three building stocks yields an annual building cooling energy consumption of about 690.7 GWh/year at the district level. In Figure 18, the daily variation of the cooling energy demand for the district is presented along with the daily average dry-bulb and wet-bulb temperatures for a year to exhibit the same correlation between the cooling energy demand and the outdoor climatic conditions as in the results for each individual archetype in Figure 13.

Table 8. Aggregated district level cooling energy consumption.

Building Type:	**Commercial**	**Residential**	**Mixed-Use**
Simulated cooling EUI (kWh/m^2/year)	209.6	225.8	219.7
Buildings' total floor area in the district (m^2)	1,215,545	600,174	1,367,492
Annual cooling energy demand (kWh/year)	254,778,232	135,519,289.2	300,437,992.4

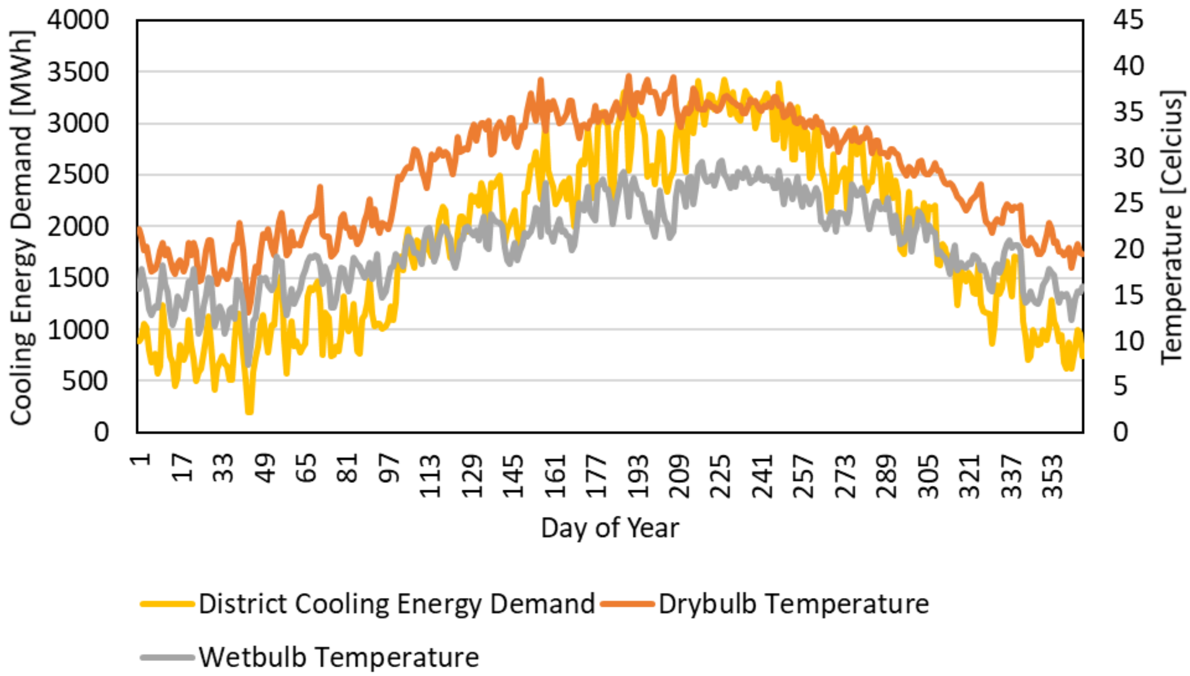


Figure 18. Annual cooling energy demand profile of the Marina District.

4.1.6. Summary

Section 4.1 presents a case study using an archetype-based approach for urban-scale energy prediction. This approach involves characterizing building archetypes to represent the three main building typologies of the district. It includes simulating the cooling load profiles to identify the main components of the cooling load at different temporal resolutions, which helps in identifying energy-saving potentials. The simulated results are validated using actual measurements and available literature to ensure the accuracy of the approach and the validity of the assumptions and inputs before aggregating the results to the district level.

The main takeaways from this section are:

- Based on the cooling load simulation results, the building cooling demand heavily relies on outdoor climatic conditions, i.e., high cooling energy demands are observed in the hot summer season, whereas the transition and winter periods show mild and low cooling demands, respectively.

- The monthly cooling load breakdown analysis indicates that the heat gained from ventilation, infiltration, and convection represents the largest weather-dependent contributors to the cooling load.
- Increasing building resilience to the outdoor weather in regions with extreme climatic conditions can yield substantial energy savings.
- The moderate winter temperatures of the region under study would allow the potential use of passive cooling strategies to minimize cooling energy needs in the winter and transition periods.

In section 4.2. The developed archetypes are expanded to include other geometrical variations present in the case study district. These archetypes serve as the initial step for generating the data needed to develop the surrogate ML models.

4.2. Surrogate ML models for urban cooling energy consumption predictions

This section describes the development of surrogate ML models for predicting the cooling EUI for the three building types in the case study district. After validating the performance of the three initial archetypes, the archetypes are expanded to include the geometrical variations identified in the district, resulting in a total of 36 archetypes (12 for each building type) as described in the methodology section. These archetypes are then used in parametric simulations with predefined input parameters to generate the necessary datasets for developing the ML models. The parametric simulations consider building thermo-physical characteristics, operational parameters, and occupant-related inputs, incorporating the influence of building geometry as well. Using these datasets, (ANNs are trained, tested, and optimized for predicting building cooling EUI. The impact of the training dataset size is investigated to determine the appropriate size for each model. Once finalized, the models are integrated into a user-friendly interface, allowing for predictions of building monthly cooling EUI for every building in the district, which are then aggregated to the district level. The developed models closely match the results from physics-based simulations but require only a fraction of the time and effort for predictions once trained. The results are discussed in the following sections.

4.2.1. Surrogate ML model development and performance evaluation

4.2.2. Evaluating training sample size effect on ML model performance

The ANNs were developed using the ‘sklearn.neaural_networks’ module from the ‘scikit-learn’ Python library. After data pre-processing, five MLP models, each having a single hidden layer with 12 neurons, were trained for each building typology using training sets of different sizes. The models were trained until either the loss or score stabilized or when the maximum number of epochs, set at 300, was reached. Each model was then evaluated using both the corresponding test and validation sets. Initially, the model's performance was visually assessed by plotting predicted outputs against ground truth values. The predicted versus true values were plotted for both the test and validation sets for the residential, commercial, and mixed-use models, as shown in Figures 19, 20 and 21 respectively. Visual evaluation of these plots indicates that all models produce accurate predictions, with predicted values closely aligning with the ground truth values, as shown by the higher density of plotted points along the line y = x. Additionally, model performance appears to improve proportionally with increased training set size. Subsequently, model performance was quantitatively evaluated using the selected performance metrics (R^2 and RMSE). The performance metrics for both the test and validation sets were averaged for each model to provide a more reliable comparison of model performance concerning training sample size, as shown in Figure 22.

The change in RMSE was used as an indicator to determine the ideal sample size for each model, as discussed and presented in Table 9. Using a 3% decrease in RMSE between successive sample sizes as a cutoff, the ideal sample sizes were identified as follows: 390 for the residential model, 650 for the commercial model, and 520 for the mixed-use model. It can be noticed that the predictions of the residential models match the true values more closely, even though they require less data for training compared to the commercial models. The mixed-use models fall in between these two in terms of data requirements and prediction accuracy. This disparity can be mainly attributed to the ranges used in the parametric simulations for generating the training data. There is a wide variation in the range used for defining the occupant-related definitions for the commercial archetypes, particularly the occupant density and equipment power density. In contrast, the residential archetypes have much narrower ranges, which seems to enhance the learning ability of the ML models, even with a smaller amount of training data.

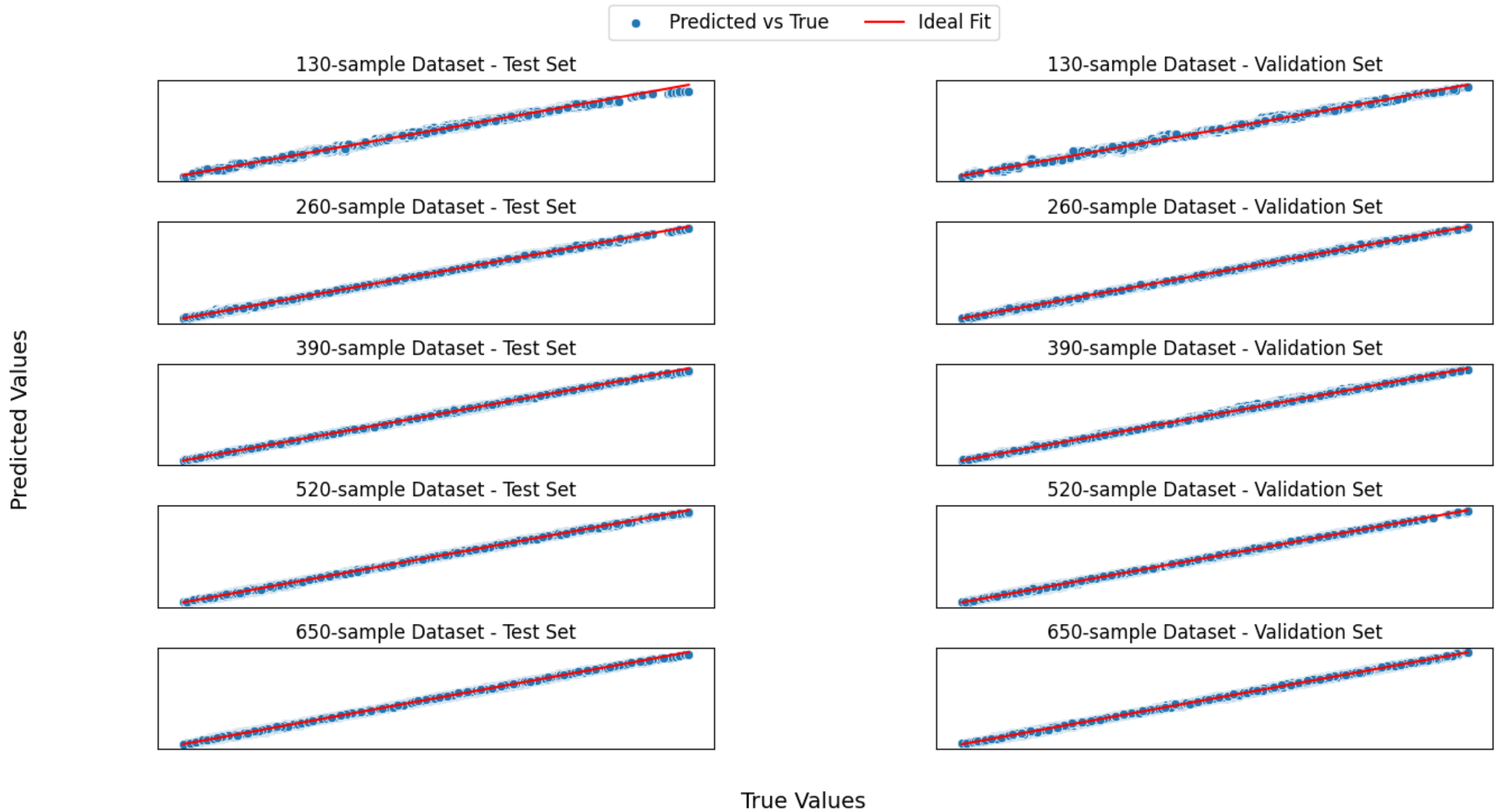


Figure 19. Predicted vs true values using both test and validation sets for residential models

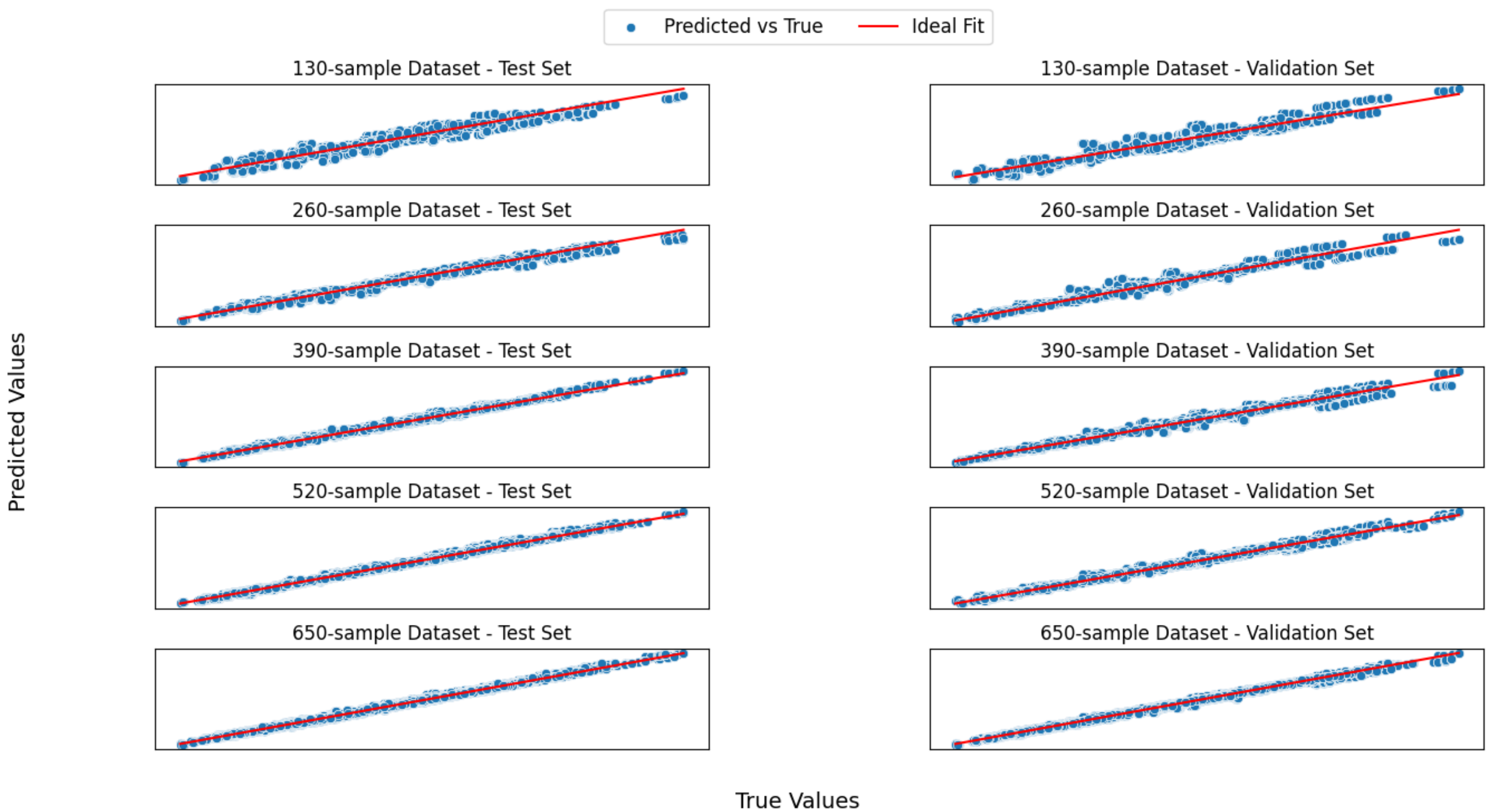


Figure 20. Predicted vs true values using both test and validation sets for commercial models

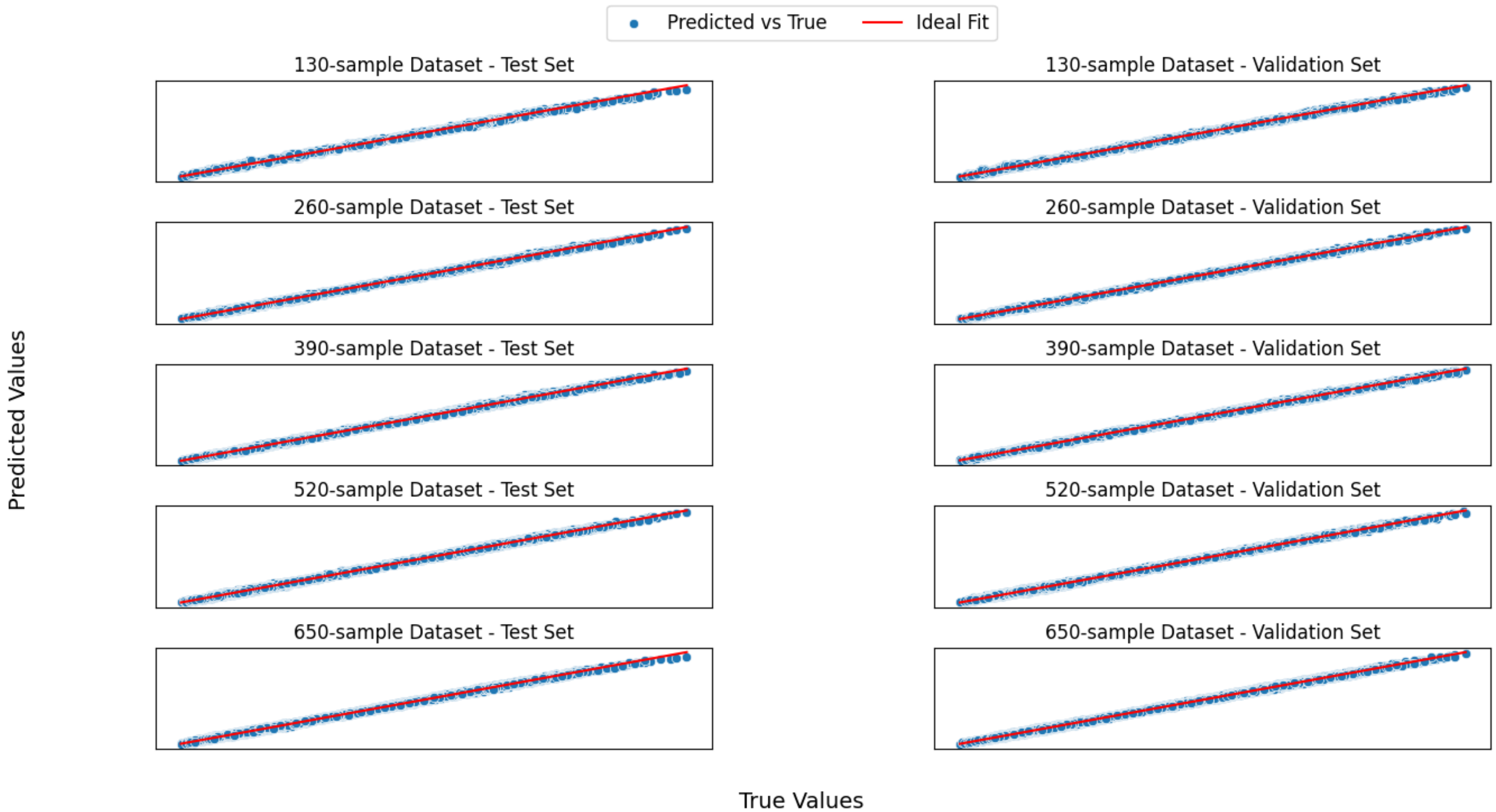


Figure 21. Predicted vs true values using both test and validation sets for mixed-use models

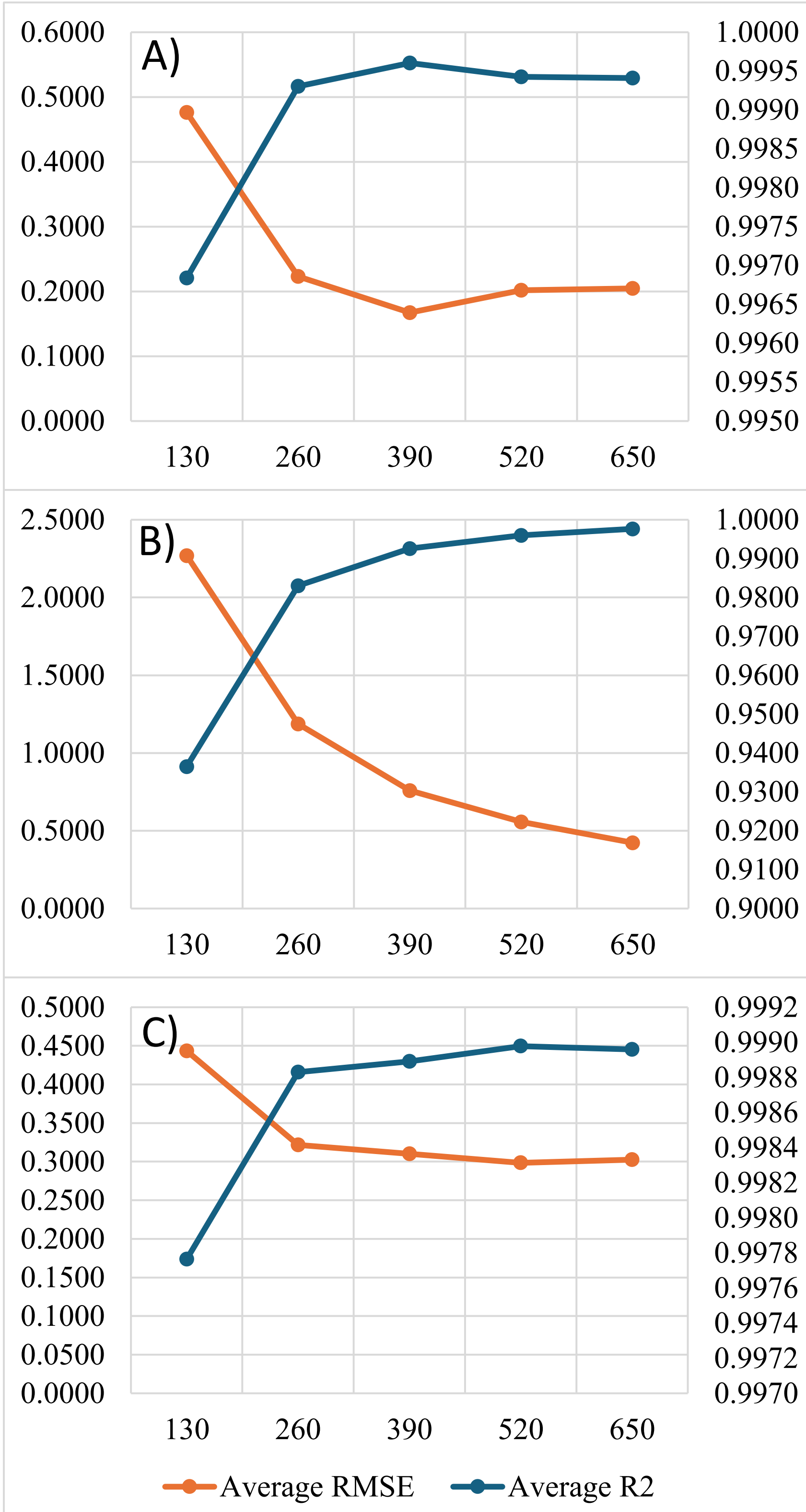


Figure 22. Average R^2 & RMSE for test and validation sets using different sample sizes for a) residential, b) commercial, and c) mixed-Use models.

Table 9. Change in average RMSE for test and validation sets between models of successive sample size.

Decrease in RMSE		Residential	Commercial	Mixed-Use
From	**To**			
130 samples	**260 samples**	53.15 %	47.73 %	27.50 %
260 samples	**390 samples**	24.98 %	35.99 %	3.48 %
390 samples	**520 samples**	**-20.69 %**	26.56 %	3.82 %
520 samples	**650 samples**	-1.25 %	24.13 %	**-1.36 %**

4.2.3. Surrogate ML model hyperparameter tuning

Following the selection of the appropriate sample size for each building typology model, a full grid search using the corresponding separate validation sets is performed to optimize the hyperparameters of each model. The chosen hyperparameters for each model are shown in Table 10.

Table 10. Hyperparameters of the ML models

Building Typology	'hidden_layer_sizes'	'activation'	'solver'	'alpha'	'learning_rate'	'batch_size'
Residential	(35,)	‘tanh’	‘adam’	0.001	‘constant’	256
Commercial	(35,)	‘tanh’	'sgd'	0.0001	‘adaptive’	64
Mixed-Use	(35,)	‘tanh’	'sgd'	0.0001	‘adaptive’	64

4.2.4. Final ML model training and performance evaluation

After defining the hyperparameters, the three models undergo a final training process using the optimized settings, incorporating an early-stop mechanism to prevent overfitting. The performance evaluation before and after hyperparameter tuning and the early-stop mechanism is presented in Table 11. While the R2 score shows a modest improvement of less than 1% for all three models, the RMSE benefits substantially from the hyperparameter tuning. Specifically, there are approximate improvements of 37%, 59%, and 56% in RMSE for the residential, commercial, and mixed-use models, respectively. These significant enhancements underscore the value and importance of hyperparameter tuning in improving the performance of ANNs. All three final models exhibit exceptionally high R2 values, which are very close to 1, indicating that the models can explain nearly all the variance in the dependent variable based on the independent variables. This suggests that the models have an excellent fit to the data. Additionally, the RMSE remains below 0.17 kWh/m^2/month for all three models, suggesting that the predictions are quite accurate. The residuals of the models are plotted as shown in Figures 23, 24, and 25. Residuals play a crucial role in assessing the quality of ML regression models. If the expected value of residuals deviates significantly from 0, it suggests that the model may exhibit systematic bias, either over-predicting or under-predicting the target variable. Furthermore, it is important to ensure that residuals are normally distributed and homoscedastic, meaning their variance remains constant over time. In contrast, heteroscedastic residuals imply that the model's predictive power varies across different sections of the data, potentially impacting the reliability of the model's predictions. For all three models, the residuals are normally distributed around a mean of 0, indicating their predictive quality.

Table 11. ML model performance before and after fine-tuning on the test set.

Building Typology	Metric	Before Tuning	After Tuning
Residential	R^2	0.99964	0.99985
	RMSE	0.1584	0.1003
Commercial	R^2	0.99778	0.99962
	RMSE	0.4118	0.1696
Mixed-Use	R^2	0.99897	0.99980
	RMSE	0.2984	0.1314

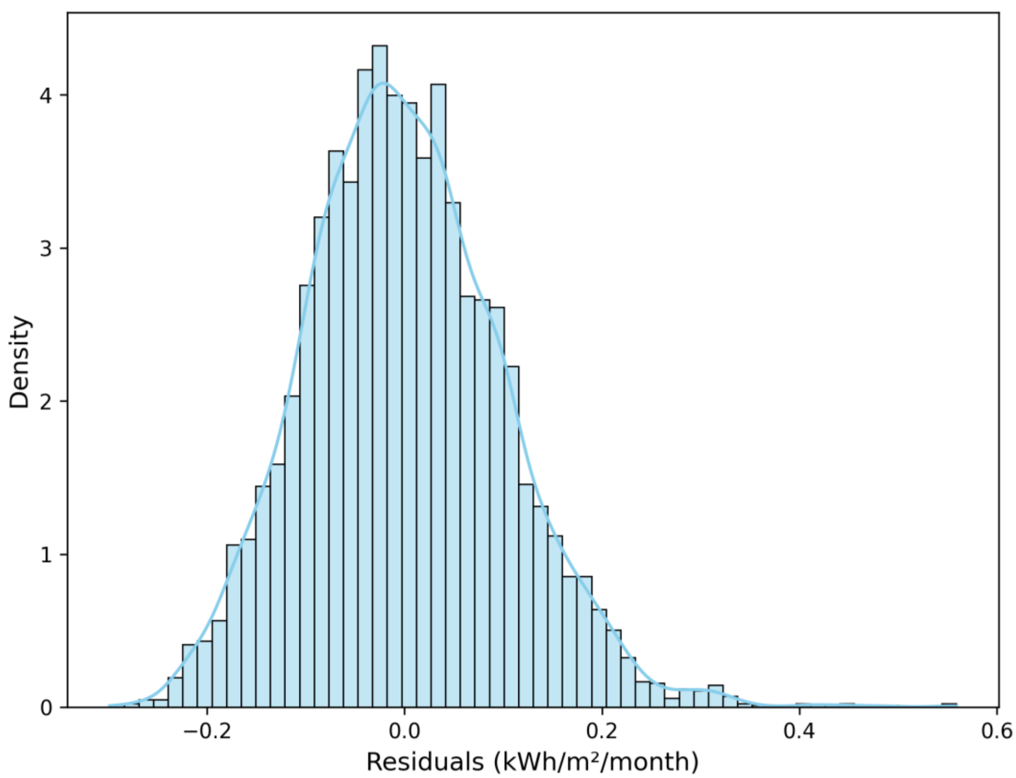


Figure 23. Residual plots for the final residential ML model.

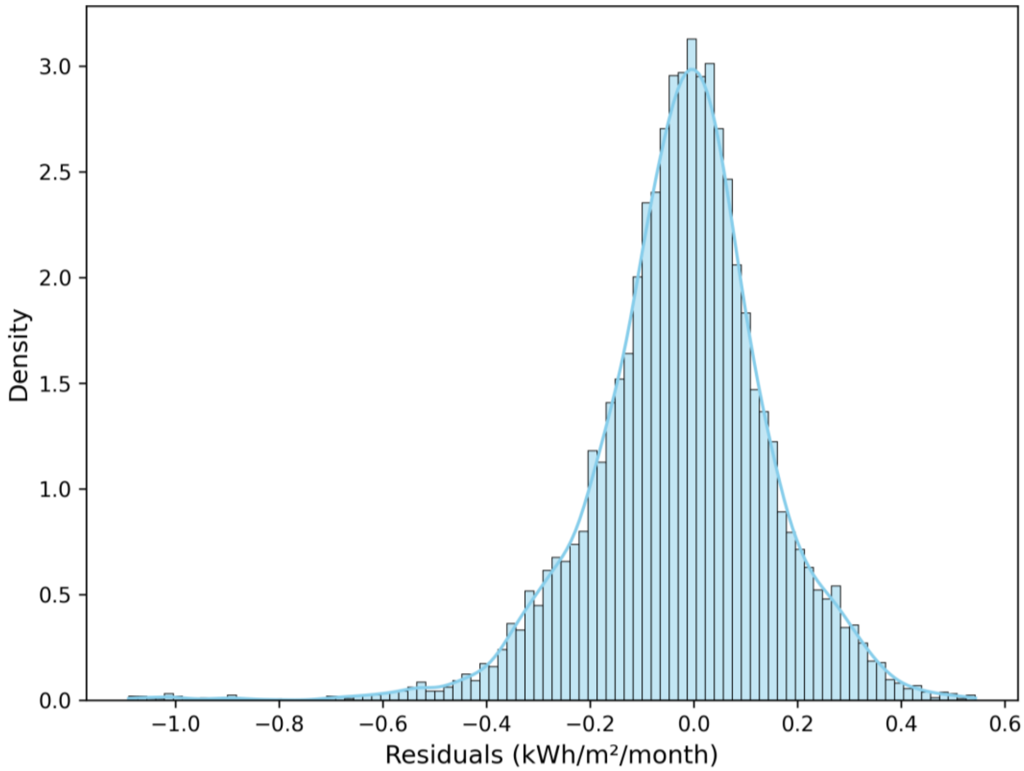


Figure 24. Residual plots for the final commercial ML model.

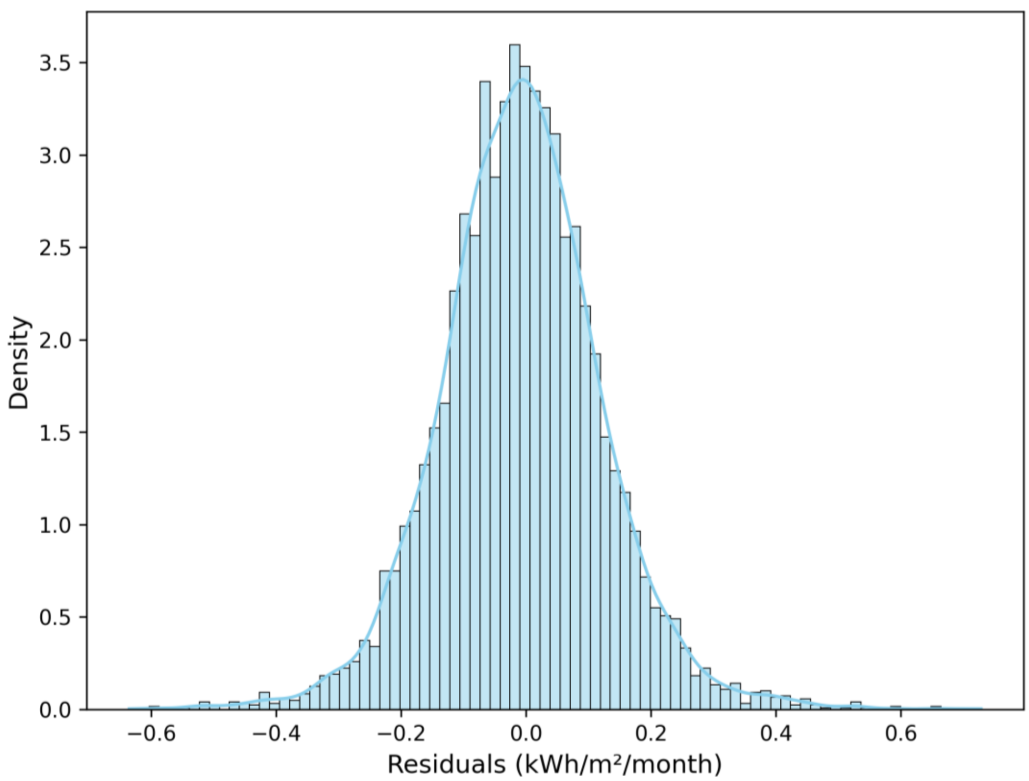


Figure 25. Residual plots for the final mixed-use ML models.

4.2.5. Decision support environment deployment for monthly cooling EUI predictions

The building information for this study was derived from the Marina District master plan, as depicted in Figure 26. In this master plan, residential buildings are represented by a light yellow color, commercial buildings are in red, and mixed-use buildings are in orange. Key information such as the number of floors, aspect ratio, and gross floor area (GFA) for each building was obtained from the master plan. A total of 107 high-rise buildings were identified, covering the three main building typologies in the district: 29 residential buildings, 37 commercial buildings, and 41 mixed-use buildings. The geometric characteristics of these buildings are detailed in Appendix 2. The decision support environment was then utilized to predict both the minimum and maximum monthly cooling EUIs for each building in the district. These EUIs were subsequently multiplied by the GFA of each building to obtain the monthly cooling energy consumption. For all predictions, a setpoint cooling temperature of 23°C, which is the standard cooling temperature for occupied spaces in Qatar, was used. Depending on whether the minimum or maximum cooling EUI was being predicted, either the upper limit or the lower limit of the parameters defined in Table 5 was used, taking into account the nature of each parameter. This approach ensured that the predictions accurately reflected the range of possible cooling energy demands for the buildings in the district. The minimum and maximum cooling energy consumption for the residential buildings are presented in Figures 27 and 28, for the commercial buildings in Figures 29 and 30, and for the mixed-use buildings in Figures 31 and 32, respectively. Section 4.2 evaluated the application of surrogate ML models to predict monthly cooling EUIs for residential, commercial, and mixed-use buildings at the district level.

Figure 26. Marina District master plan

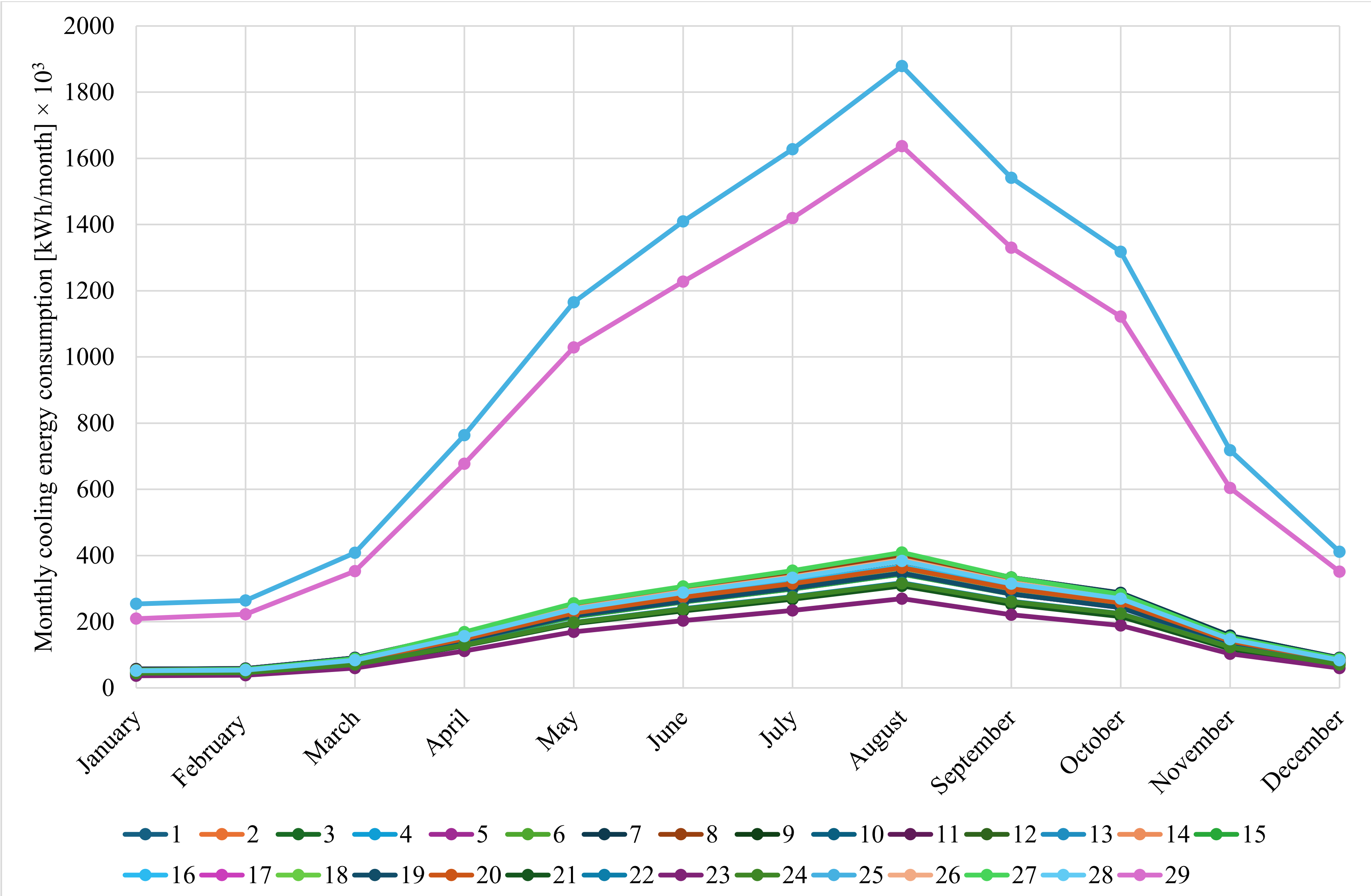


Figure 27. Minimum monthly cooling energy consumption for residential buildings.

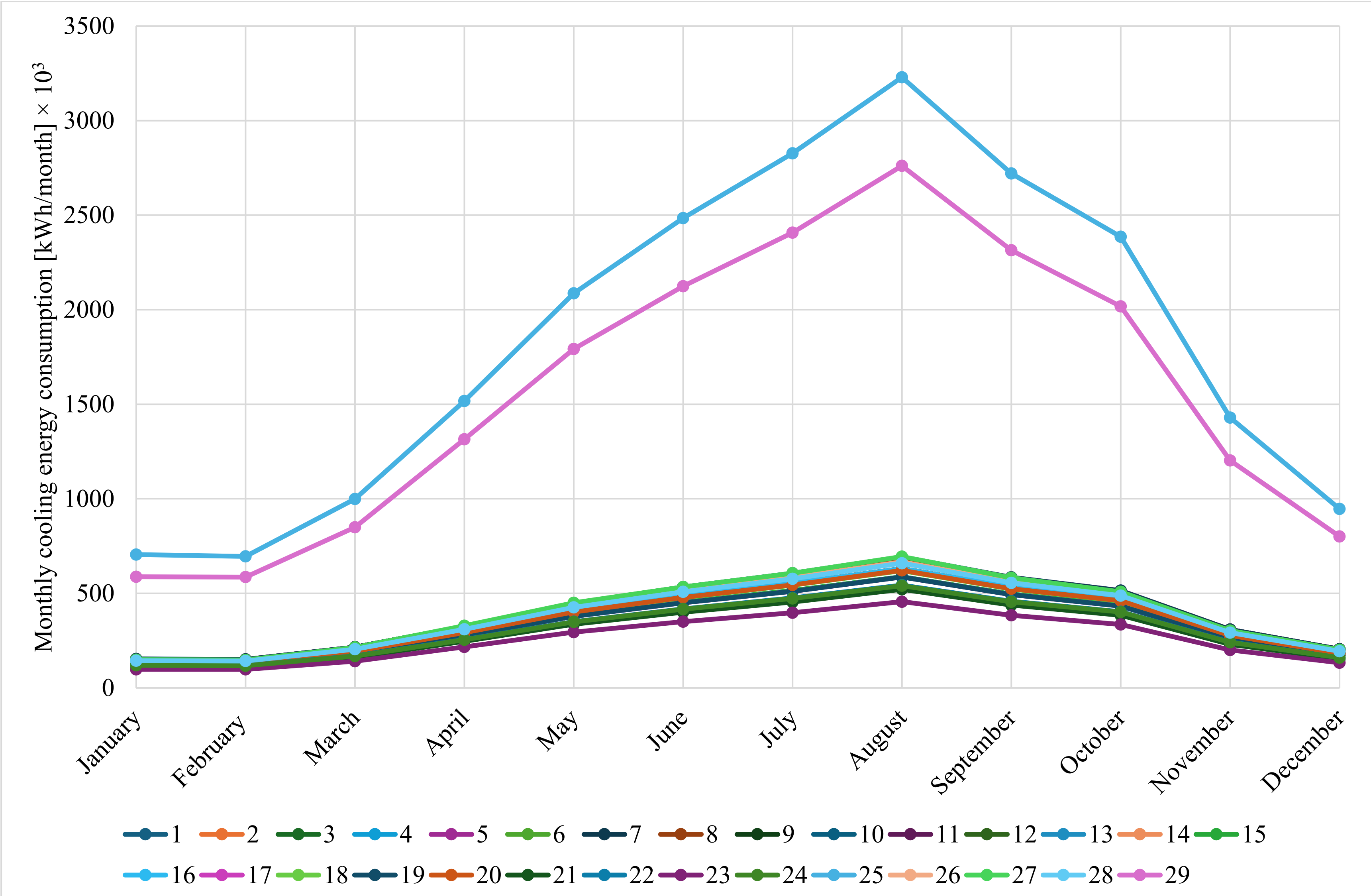


Figure 28. Maximum monthly cooling energy consumption for residential buildings.

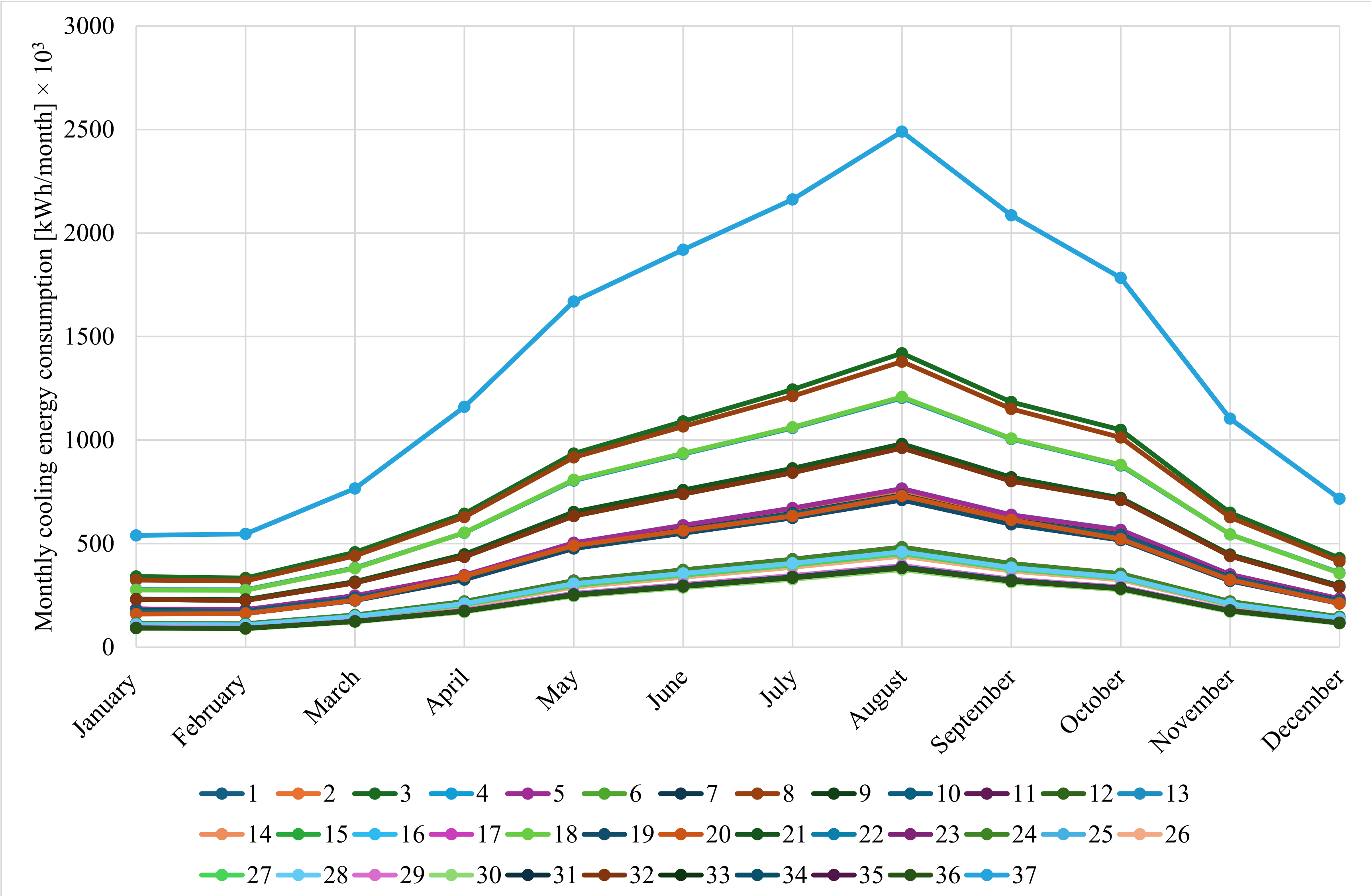


Figure 29. Minimum monthly cooling energy consumption for commercial buildings.

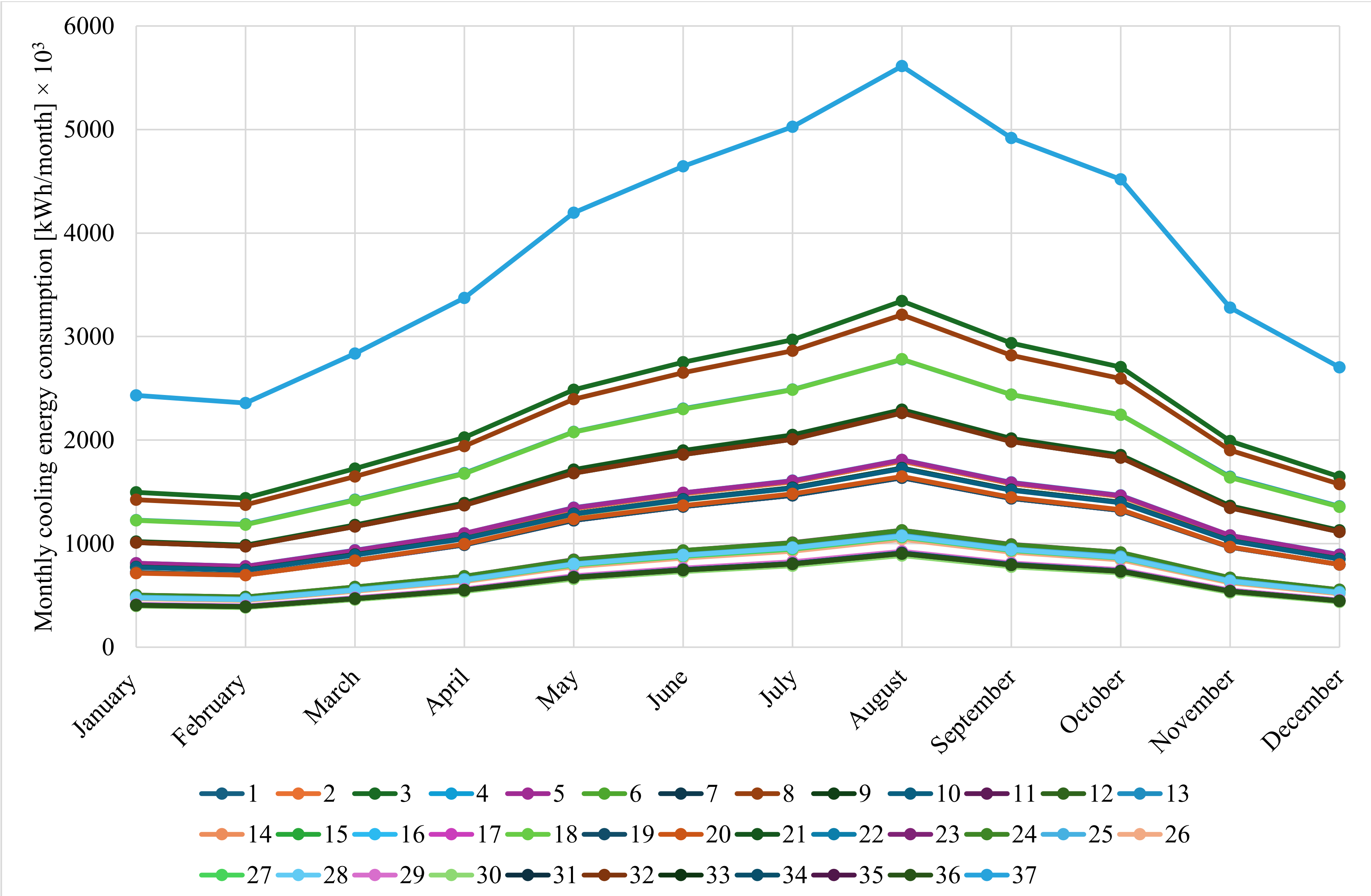


Figure 30. Maximum monthly cooling energy consumption for commercial buildings.

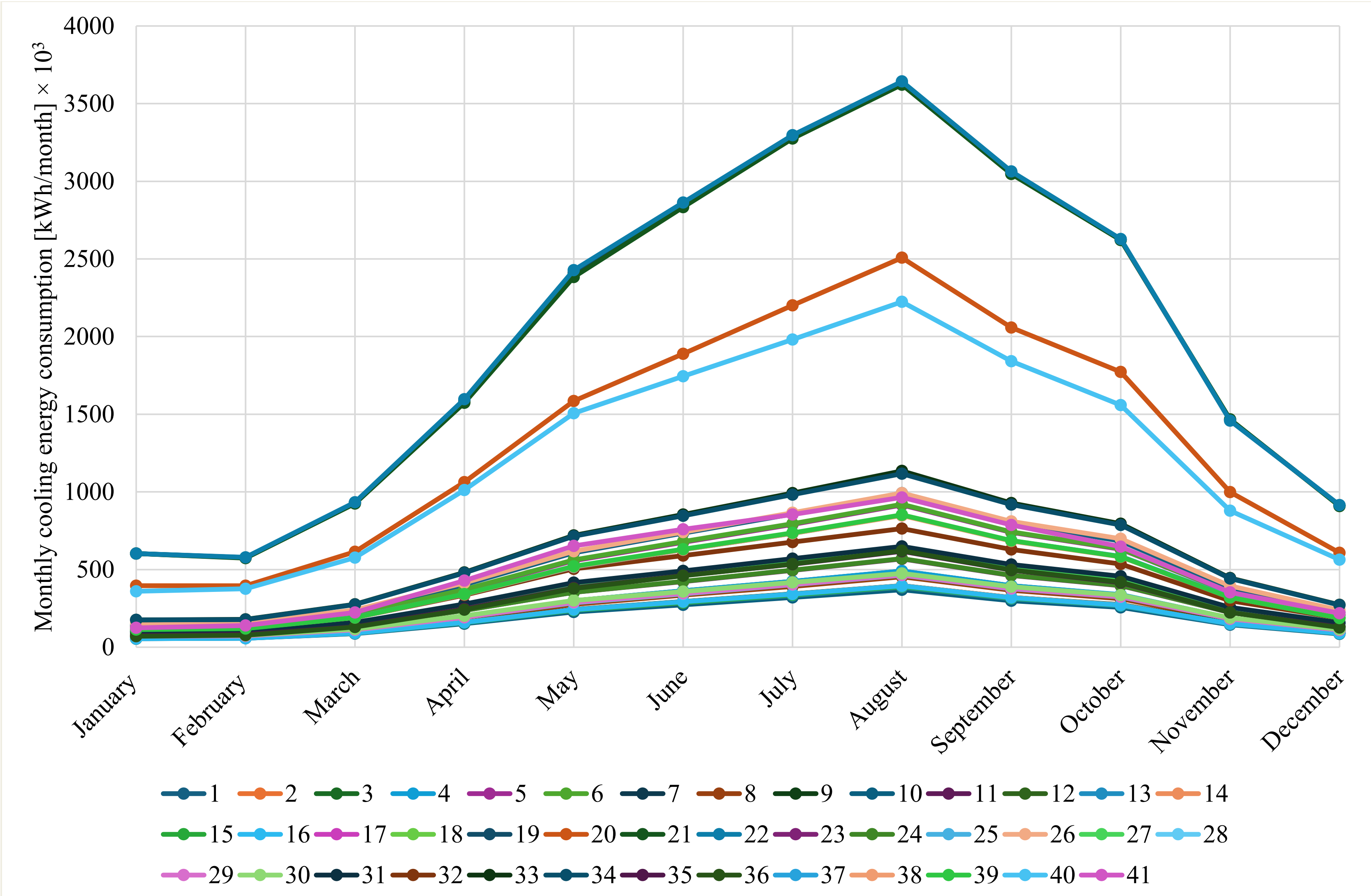


Figure 31. Minimum monthly cooling energy consumption for mixed-use buildings.

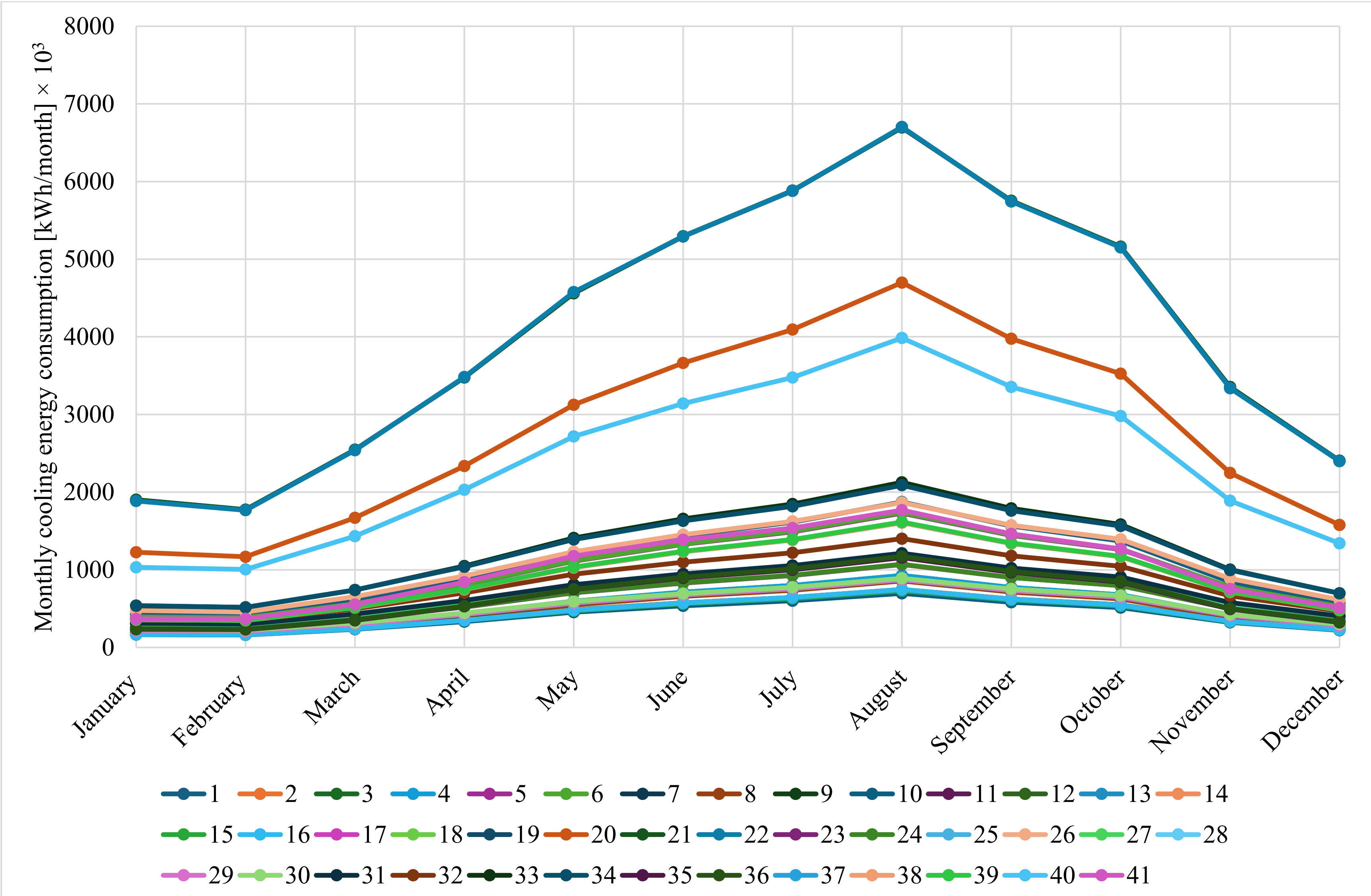


Figure 32. Maximum monthly cooling energy consumption for mixed-use buildings.

4.2.6. District-level cooling energy consumption

By aggregating the results for each building, the total monthly cooling energy consumption for the district can be calculated, as shown in Figure 33. Instead of relying on average values, the district's cooling energy consumption is presented as a range, accounting for the impacts of input uncertainty. The district's maximum cooling energy consumption can surpass 170 GWh in August. A narrower range of cooling energy consumption is observed in the winter months compared to the summer months. This significant uncertainty is mainly due to the wide ranges of input parameters used in the data generation process. To reduce this uncertainty in surrogate ML models, further efforts are needed to refine the input parameter distributions and reduce their variability. By obtaining more precise information regarding the district's design and operational parameters, the decision support environment can provide quick and accurate estimations of cooling energy consumption. This significantly aids in the sizing and operation of DCP, ensuring it is adequately equipped to meet the cooling demands of the district efficiently.

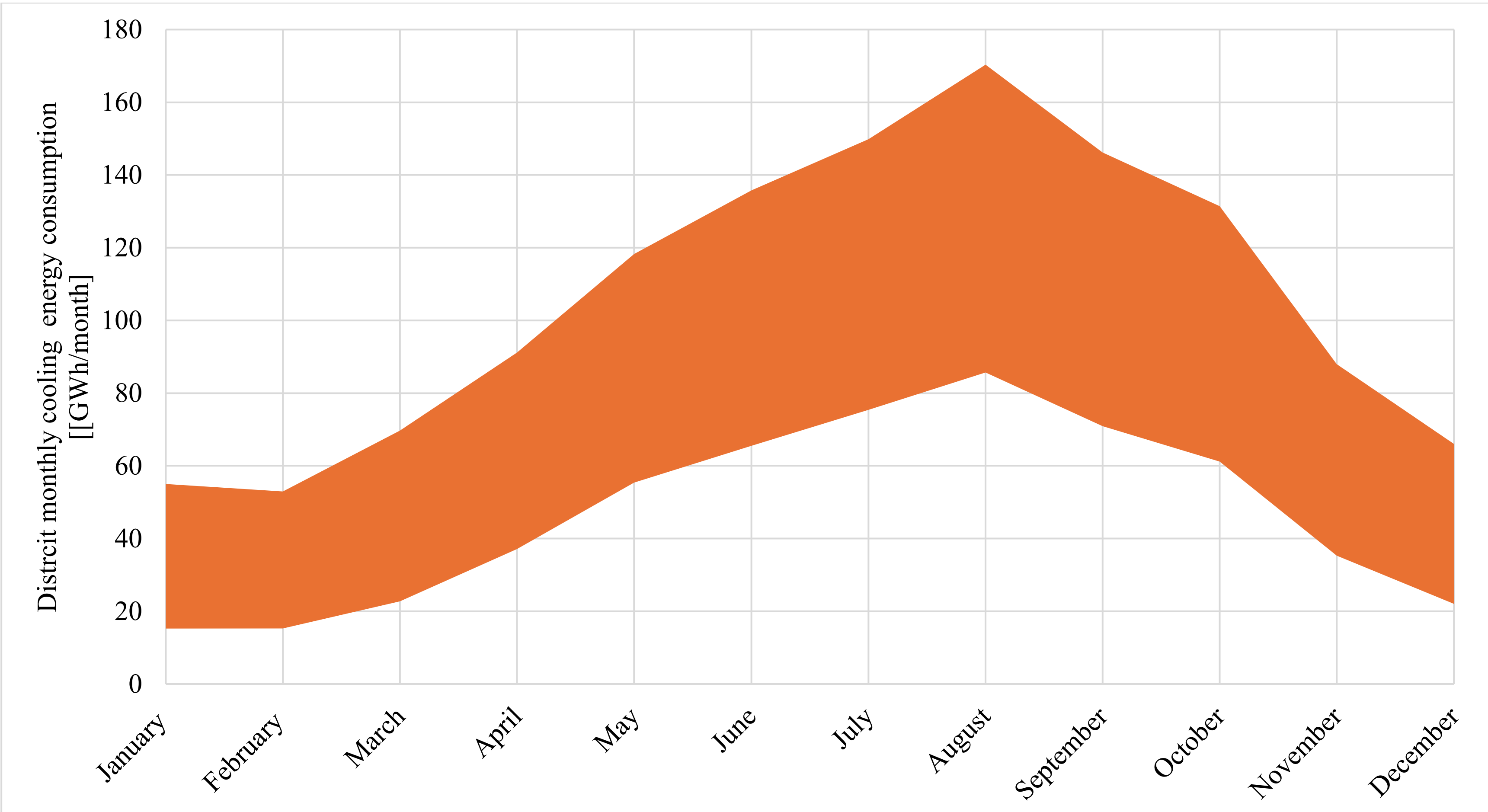


Figure 33. Range of total cooling energy consumption of the Marina District

4.2.7. Summary

Section 4.2 examined the application of surrogate ML models to predict the monthly cooling EUIs for residential, commercial, and mixed-use high-rise buildings at the district level. Parametric simulations were conducted using building physics-based archetype models to create the required datasets for developing the ML models. The effect of training dataset size on ML model performance was analyzed for each building typology to determine the appropriate size for each building model. The input parameters included climate factors (outdoor dry bulb temperature and relative humidity), building geometry parameters (floor number and aspect ratio), thermo-physical characteristics (window, wall, and roof U-values and window SHGC), building operational parameters (temperature setpoints, infiltration, and ventilation rates), occupant-related definitions (occupancy, lighting, and electric equipment densities), and occupant-related schedules (occupancy, lighting, and electric equipment utilization percentages). The developed models exhibit superior performance with R^2 values close to 1 and RMSE less than 0.17 kWh/m^2/month on unseen data, highlighting the potential of ML surrogate models in optimizing both the design and operation of building and district energy systems. These models offer a more comprehensive depiction of the factors impacting building energy performance by accounting for building design, building operation, and occupant behavior aspects. The proposed models were also integrated within a user-friendly decision support interface to facilitate their use. This decision support interface was then used to predict the minimum and maximum monthly cooling EUI profiles for each building in the district within seconds, before aggregating the results to the district level.

.

CHAPTER 5

5. Conclusions and future work

5.1. Conclusion

This study proposed a methodology to predict district-level cooling energy consumption for high-rise buildings in a hot and arid climate. Initially, three high-rise building archetypes were developed to represent the main building typologies in the case study district. Their cooling load profiles were simulated, validated with measured data, and analyzed to identify energy-saving potential. The results were then aggregated to the district level. This step ensures the accuracy of the inputs used for characterizing the archetypes and demonstrates a common method for urban-scale building energy analysis. These archetypes serve as the foundation for creating surrogate ML models. The three initial archetypes were expanded to include different geometrical variations prevalent in the case study district. In total, 12 building archetypes per building typology, amounting to 36 archetypes, were developed to represent the geometrical variations in the case study district. Parametric simulations were then performed to create the required datasets for ML model development. The ML models were subsequently trained, tested, and optimized to predict cooling EUIs for each of the considered building types. The input parameters for the surrogate ML models included climate factors (outdoor dry bulb temperature and relative humidity), building geometry parameters (floor number and aspect ratio), thermo-physical characteristics (window, wall, and roof U-values and window SHGC), building operational parameters (temperature setpoints, infiltration, and ventilation rates), occupant-related definitions (occupancy, lighting, and electric equipment densities), and occupant-related schedules (occupancy, lighting, and electric equipment utilization percentages). The developed models consider both building design parameters and building operation-related parameters, incorporating occupant-related factors, as independent variables for making predictions, thus providing a comprehensive representation of the variables affecting building energy performance. The developed models exhibit superior performance, with R^2 values close to 1 and RMSE less than 0.17 kWh/m²/month on unseen data, highlighting the potential of ML surrogate models in optimizing both the design and operation of building and district energy systems. The trained ML models are then integrated with a user-friendly decision support environment (Excel sheet) to facilitate their usability. This decision support environment was used to predict the minimum and maximum monthly cooling EUI profile for every building in the district in a matter of seconds before aggregating the results to obtain the district-level profile. The tool also provides a means of investigating various design, operation, and retrofit scenarios at a fraction of the computational cost, compared to physics-based UBEM approaches. Incorporating inputs related to occupant behavior and building operation explores the potential for building energy predictions considering users' activities and characteristics, allowing for a more accurate representation of the diversity among buildings of the same type or physical characteristics.

The main contributions of this work are as follows:

- Developing and validating building archetype models to represent the high-rise building stock in hot and arid climates.
- Analyzing the cooling load profiles of the developed archetypes and investigating the energy-saving potential of high-rise buildings in such climates.
- Proposing an overall methodology for developing surrogate ML models to predict monthly cooling EUIs for high-rise residential, commercial, and mixed-use buildings in an urban setting

with limited data scenarios.

- Considering both building design parameters and building operation-related parameters, incorporating occupant-related factors as independent variables for making predictions using surrogate ML models, thus providing a comprehensive representation of the variables affecting building energy performance.
- Integrating the trained ML models with a user-friendly interface that facilitates straightforward predictions of cooling EUIs and allows for the testing of different design alternatives, thereby enhancing the decision-making process.

5.2. Limitations and future work

There exists several limitations that were encountered in this study:

- For validating the three initial archetypes, data was available from only one residential and one mixed-use building, with no data available for commercial buildings. Additionally, having data for extended periods and from multiple buildings per typology would enhance the validation process and enable better calibration of the input ranges.
- Only rectangular-shaped buildings were considered in this case study, excluding non-rectangular building shapes. Other building shape variations should be considered in future works to present more comprehensive models capable of representing the high-rise building stock in that climatic region.
- Due to the lack of specific details regarding HVAC system specifications, an ideal load system was assumed in all EnergyPlus simulations. In future research, different HVAC configurations should be considered to enable a more comprehensive analysis.
- Urban microclimate effects and building-to-building interactions were not considered in this

study.

- Occupant behavior schedules were based on ASHRAE standards due to the absence of region-specific schedules or data. Incorporating region-specific occupant behavior data would further enhance the reliability of the models.
- The wide ranges used for some input parameters increased the uncertainty in predictions and affected model performance. Efforts to calibrate input parameter ranges or conduct wide-scale data collection campaigns would reduce this uncertainty and enhance model accuracy and reliability.

6. References

[1] "Energy Technology Perspectives 2017 – Analysis - IEA." Accessed: Jan. 03, 2023. [Online]. Available: https://www.iea.org/reports/energy-technology-perspectives-2017

[2] M. C. Peel, B. L. Finlayson, and T. A. McMahon, "Updated world map of the Köppen-Geiger climate classification," *Hydrol Earth Syst Sci*, vol. 11, no. 5, pp. 1633–1644, 2007, doi: 10.5194/HESS-11-1633-2007.

[3] N. Ayoub, F. Musharavati, S. Pokharel, and H. A. Gabbar, "Energy consumption and conservation practices in Qatar—A case study of a hotel building," *Energy Build*, vol. 84, pp. 55–69, Dec. 2014, doi: 10.1016/J.ENBUILD.2014.07.050.

[4] M. I. Fahkroo, N. Al-Awainati, F. Musharavati, S. Pokherel, and H. A. Gabbar, "Operations optimization towards high performance cooling in commercial buildings," *IEEE International Conference on Smart Energy Grid Engineering, SEGE 2013*, 2013, doi: 10.1109/SEGE.2013.6707930.

[5] P. A. Østergaard *et al.*, "The four generations of district cooling - A categorization of the development in district cooling from origin to future prospect," *Energy*, vol. 253, p. 124098, Aug. 2022, doi: 10.1016/J.ENERGY.2022.124098.

[6] "Qatar General Electricity & Water Corporation 'KAHRAMAA' Statistic Report 2019," 2020.

[7] C. F. Reinhart and C. Cerezo Davila, "Urban building energy modeling – A review of a nascent field," *Build Environ*, vol. 97, pp. 196–202, Feb. 2016, doi: 10.1016/J.BUILDENV.2015.12.001.

[8] Y. Chen, M. Guo, Z. Chen, Z. Chen, and Y. Ji, "Physical energy and data-driven models in building energy prediction: A review," *Energy Reports*, vol. 8, pp. 2656–2671, Nov. 2022, doi: 10.1016/J.EGYR.2022.01.162.

[9] M. Bourdeau, X. qiang Zhai, E. Nefzaoui, X. Guo, and P. Chatellier, "Modeling and forecasting building energy consumption: A review of data-driven techniques," *Sustain Cities Soc*, vol. 48, p. 101533, Jul. 2019, doi: 10.1016/J.SCS.2019.101533.

[10] J. Yu, W. S. Chang, and Y. Dong, "Building Energy Prediction Models and Related Uncertainties: A Review," *Buildings 2022, Vol. 12, Page 1284*, vol. 12, no. 8, p. 1284, Aug. 2022, doi: 10.3390/BUILDINGS12081284.

[11] G. Happle, J. A. Fonseca, and A. Schlueter, "A review on occupant behavior in urban

building energy models," *Energy Build*, vol. 174, pp. 276–292, Sep. 2018, doi: 10.1016/J.ENBUILD.2018.06.030.

[12] F. Johari, G. Peronato, P. Sadeghian, X. Zhao, and J. Widén, "Urban building energy modeling: State of the art and future prospects," *Renewable and Sustainable Energy Reviews*, vol. 128, p. 109902, Aug. 2020, doi: 10.1016/J.RSER.2020.109902.

[13] O. Ahmed, N. Sezer, M. Ouf, L. (Leon) Wang, and I. G. Hassan, "State-of-the-art review of occupant behavior modeling and implementation in building performance simulation," *Renewable and Sustainable Energy Reviews*, vol. 185, p. 113558, Oct. 2023, doi: 10.1016/J.RSER.2023.113558.

[14] İ. Dinçer and C. Zamfirescu, "Sustainable energy systems and applications," *Sustainable Energy Systems and Applications*, pp. 1–816, Jan. 2012, doi: 10.1007/978-0-387-95861-3/COVER.

[15] "District Cooling Best Practice Guide - International District Energy Association." Accessed: Jan. 04, 2023. [Online]. Available: https://www.districtenergy.org/resources/publications/district-cooling-best-practices

[16] V. Eveloy and D. S. Ayou, "Sustainable District Cooling Systems: Status, Challenges, and Future Opportunities, with Emphasis on Cooling-Dominated Regions," *Energies 2019, Vol. 12, Page 235*, vol. 12, no. 2, p. 235, Jan. 2019, doi: 10.3390/EN12020235.

[17] W. Gang, S. Wang, F. Xiao, and D. C. Gao, "District cooling systems: Technology integration, system optimization, challenges and opportunities for applications," *Renewable and Sustainable Energy Reviews*, vol. 53, pp. 253–264, Jan. 2016, doi: 10.1016/J.RSER.2015.08.051.

[18] Y. Shimoda, T. Nagota, N. Isayama, and M. Mizuno, "Verification of energy efficiency of district heating and cooling system by simulation considering design and operation parameters," *Build Environ*, vol. 43, no. 4, pp. 569–577, Apr. 2008, doi: 10.1016/J.BUILDENV.2006.03.017.

[19] A. Perdichizzi, G. Barigozzi, G. Franchini, and S. Ravelli, "Peak shaving strategy through a solar combined cooling and power system in remote hot climate areas," *Appl Energy*, vol. 143, pp. 154–163, Apr. 2015, doi: 10.1016/J.APENERGY.2015.01.030.

[20] C. Marugán-Cruz, S. Sánchez-Delgado, M. R. Rodríguez-Sánchez, M. Venegas, and D. Santana, "District cooling network connected to a solar power tower," *Appl Therm Eng*,

vol. 79, pp. 174–183, Mar. 2015, doi: 10.1016/J.APPLTHERMALENG.2015.01.032.
[21] M. Ameri and Z. Besharati, “Optimal design and operation of district heating and cooling networks with CCHP systems in a residential complex,” *Energy Build*, vol. 110, pp. 135–148, Jan. 2016, doi: 10.1016/J.ENBUILD.2015.10.050.
[22] C. F. Reinhart and C. Cerezo Davila, “Urban building energy modeling – A review of a nascent field,” *Build Environ*, vol. 97, pp. 196–202, Feb. 2016, doi: 10.1016/J.BUILDENV.2015.12.001.
[23] T. Hong, Y. Chen, X. Luo, N. Luo, and S. H. Lee, “Ten questions on urban building energy modeling,” *Build Environ*, vol. 168, p. 106508, Jan. 2020, doi: 10.1016/J.BUILDENV.2019.106508.
[24] Y. Q. Ang, Z. M. Berzolla, and C. F. Reinhart, “From concept to application: A review of use cases in urban building energy modeling,” *Appl Energy*, vol. 279, p. 115738, Dec. 2020, doi: 10.1016/J.APENERGY.2020.115738.
[25] R. Giridharan, S. S. Y. Lau, S. Ganesan, and B. Givoni, “Urban design factors influencing heat island intensity in high-rise high-density environments of Hong Kong,” *Build Environ*, vol. 42, no. 10, pp. 3669–3684, Oct. 2007, doi: 10.1016/J.BUILDENV.2006.09.011.
[26] A. L. Pisello, J. E. Taylor, X. Xu, and F. Cotana, “Inter-building effect: Simulating the impact of a network of buildings on the accuracy of building energy performance predictions,” *Build Environ*, vol. 58, pp. 37–45, Dec. 2012, doi: 10.1016/J.BUILDENV.2012.06.017.
[27] J. A. Fonseca and A. Schlueter, “Integrated model for characterization of spatiotemporal building energy consumption patterns in neighborhoods and city districts,” *Appl Energy*, vol. 142, pp. 247–265, Mar. 2015, doi: 10.1016/J.APENERGY.2014.12.068.
[28] C. Delmastro, F. Martinsson, G. Mutani, and S. P. Corgnati, “Modeling Building Energy Demand Profiles and District Heating Networks for Low Carbon Urban Areas,” *Procedia Eng*, vol. 198, pp. 386–397, Jan. 2017, doi: 10.1016/J.PROENG.2017.07.094.
[29] O. Pasichnyi, J. Wallin, and O. Kordas, “Data-driven building archetypes for urban building energy modelling,” *Energy*, vol. 181, pp. 360–377, Aug. 2019, doi: 10.1016/J.ENERGY.2019.04.197.
[30] C. Cerezo Davila, C. F. Reinhart, and J. L. Bemis, “Modeling Boston: A workflow for the

efficient generation and maintenance of urban building energy models from existing geospatial datasets," *Energy*, vol. 117, pp. 237–250, Dec. 2016, doi: 10.1016/J.ENERGY.2016.10.057.

[31] S. Nagpal, C. Mueller, A. Aijazi, and C. F. Reinhart, "A methodology for auto-calibrating urban building energy models using surrogate modeling techniques," *J Build Perform Simul*, vol. 12, no. 1, pp. 1–16, Jan. 2019, doi: 10.1080/19401493.2018.1457722.

[32] A. Nutkiewicz, Z. Yang, and R. K. Jain, "Data-driven Urban Energy Simulation (DUE-S): A framework for integrating engineering simulation and machine learning methods in a multi-scale urban energy modeling workflow," *Appl Energy*, vol. 225, pp. 1176–1189, Sep. 2018, doi: 10.1016/J.APENERGY.2018.05.023.

[33] M. H. Kristensen, R. E. Hedegaard, and S. Petersen, "Hierarchical calibration of archetypes for urban building energy modeling," *Energy Build*, vol. 175, pp. 219–234, Sep. 2018, doi: 10.1016/J.ENBUILD.2018.07.030.

[34] M. Razmara, G. R. Bharati, D. Hanover, M. Shahbakhti, S. Paudyal, and R. D. Robinett, "Building-to-grid predictive power flow control for demand response and demand flexibility programs," *Appl Energy*, vol. 203, pp. 128–141, Oct. 2017, doi: 10.1016/J.APENERGY.2017.06.040.

[35] B. Dong, Z. Li, A. Taha, and N. Gatsis, "Occupancy-based buildings-to-grid integration framework for smart and connected communities," *Appl Energy*, vol. 219, pp. 123–137, Jun. 2018, doi: 10.1016/J.APENERGY.2018.03.007.

[36] R. Yin, S. Kiliccote, and M. A. Piette, "Linking measurements and models in commercial buildings: A case study for model calibration and demand response strategy evaluation," *Energy Build*, vol. 124, pp. 222–235, Jul. 2016, doi: 10.1016/J.ENBUILD.2015.10.042.

[37] R. Yin *et al.*, "Quantifying flexibility of commercial and residential loads for demand response using setpoint changes," *Appl Energy*, vol. 177, pp. 149–164, Sep. 2016, doi: 10.1016/J.APENERGY.2016.05.090.

[38] D. Kong *et al.*, "Urban building energy modeling (UBEM): a systematic review of challenges and opportunities," *Energy Efficiency 2023 16:6*, vol. 16, no. 6, pp. 1–42, Aug. 2023, doi: 10.1007/S12053-023-10147-Z.

[39] Z. Li, J. Ma, Y. Tan, C. Guo, and X. Li, "Combining physical approaches with deep learning techniques for urban building energy modeling: A comprehensive review and

future research prospects," *Build Environ*, vol. 246, p. 110960, Dec. 2023, doi: 10.1016/J.BUILDENV.2023.110960.

[40] C. W. Lee and J. Zhong, "Top down strategy for renewable energy investment: Conceptual framework and implementation," *Renew Energy*, vol. 68, pp. 761–773, Aug. 2014, doi: 10.1016/J.RENENE.2014.03.015.

[41] D. J. Sailor and L. Lu, "A top–down methodology for developing diurnal and seasonal anthropogenic heating profiles for urban areas," *Atmos Environ*, vol. 38, no. 17, pp. 2737–2748, Jun. 2004, doi: 10.1016/J.ATMOSENV.2004.01.034.

[42] 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," *Sustain Cities Soc*, vol. 62, p. 102408, Nov. 2020, doi: 10.1016/J.SCS.2020.102408.

[43] L. G. Swan and V. I. Ugursal, "Modeling of end-use energy consumption in the residential sector: A review of modeling techniques," *Renewable and Sustainable Energy Reviews*, vol. 13, no. 8, pp. 1819–1835, Oct. 2009, doi: 10.1016/J.RSER.2008.09.033.

[44] U. Ali, M. H. Shamsi, C. Hoare, E. Mangina, and J. O'Donnell, "Review of urban building energy modeling (UBEM) approaches, methods and tools using qualitative and quantitative analysis," *Energy Build*, vol. 246, p. 111073, Sep. 2021, doi: 10.1016/J.ENBUILD.2021.111073.

[45] U. Ali *et al.*, "Urban building energy performance prediction and retrofit analysis using data-driven machine learning approach," *Energy Build*, vol. 303, p. 113768, Jan. 2024, doi: 10.1016/J.ENBUILD.2023.113768.

[46] T. Hong, S. H. Lee, M. A. Piette, Y. Chen, and M. A. Piette, "CityBES: A web-based platform to support city-scale building energy efficiency," *researchgate.netT Hong, Y Chen, SH Lee, MA PietteUrban Computing, 2016•researchgate.net*, 2016, doi: 10.1145/12345.67890.

[47] T. Vermeulen, ; J H Kämpf, and ; B Beckers, "Urban form optimization for the energy performance of buildings using Citysim," 2013, Accessed: Feb. 25, 2024. [Online]. Available: https://infoscience.epfl.ch/record/195704

[48] J. A. Fonseca, T. A. Nguyen, A. Schlueter, and F. Marechal, "City Energy Analyst (CEA): Integrated framework for analysis and optimization of building energy systems in neighborhoods and city districts," *Energy Build*, vol. 113, pp. 202–226, Feb. 2016, doi:

10.1016/J.ENBUILD.2015.11.055.

[49] C. F. Reinhart, T. Dogan, J. A. Jakubiec, T. Rakha, and A. Sang, "UMI - An urban simulation environment for building energy use, daylighting and walkability," *Proceedings of BS 2013: 13th Conference of the International Building Performance Simulation Association*, pp. 476–483, Jan. 2013, doi: 10.26868/25222708.2013.1404.

[50] R. Baetens, R. De Coninck, F. Jorissen, … D. P.-B., and undefined 2015, "Openideas-an open framework for integrated district energy simulations," *researchgate.netR Baetens, R De Coninck, F Jorissen, D Picard, L Helsen, D SaelensBuilding simulation, 2015•researchgate.net*, doi: 10.26868/25222708.2015.2243.

[51] P. Remmen, M. Lauster, M. Mans, M. Fuchs, T. Osterhage, and D. Müller, "TEASER: an open tool for urban energy modelling of building stocks," *Taylor & FrancisP Remmen, M Lauster, M Mans, M Fuchs, T Osterhage, D MüllerJournal of Building Performance Simulation, 2018•Taylor & Francis*, vol. 11, no. 1, pp. 84–98, Jan. 2018, doi: 10.1080/19401493.2017.1283539.

[52] N. Abbasabadi and J. K. Mehdi Ashayeri, "Urban energy use modeling methods and tools: A review and an outlook," *Build Environ*, vol. 161, p. 106270, Aug. 2019, doi: 10.1016/J.BUILDENV.2019.106270.

[53] Kontokosta, "Predicting building energy efficiency using New York City benchmarking data," *Proceedings of the 2012 ACEEE Summer Study on Energy Efficiency in Buildings*, 2012, Accessed: Feb. 25, 2024. [Online]. Available: https://www.aceee.org/files/proceedings/2012/data/papers/0193-000114.pdf

[54] P. Kuusela, I. Norros, R. Weiss, and T. Sorasalmi, "Practical lognormal framework for household energy consumption modeling," *Energy Build*, vol. 108, pp. 223–235, Dec. 2015, doi: 10.1016/J.ENBUILD.2015.09.008.

[55] A. Mastrucci, O. Baume, F. Stazi, and U. Leopold, "Estimating energy savings for the residential building stock of an entire city: A GIS-based statistical downscaling approach applied to Rotterdam," *Energy Build*, vol. 75, pp. 358–367, Jun. 2014, doi: 10.1016/J.ENBUILD.2014.02.032.

[56] P. Dagnely, T. Ruette, T. Tourwé, E. Tsiporkova, and C. Verhelst, "Predicting hourly energy consumption. can regression modeling improve on an autoregressive baseline?," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial*

Intelligence and Lecture Notes in Bioinformatics), vol. 9518, pp. 105–122, 2015, doi: 10.1007/978-3-319-27430-0_8/TABLES/2.

[57] A. Rahman, V. Srikumar, and A. D. Smith, “Predicting electricity consumption for commercial and residential buildings using deep recurrent neural networks,” *Appl Energy*, vol. 212, pp. 372–385, Feb. 2018, doi: 10.1016/J.APENERGY.2017.12.051.

[58] C. Robinson *et al.*, “Machine learning approaches for estimating commercial building energy consumption,” *Appl Energy*, vol. 208, pp. 889–904, Dec. 2017, doi: 10.1016/J.APENERGY.2017.09.060.

[59] Y. Zhang, B. K. Teoh, M. Wu, J. Chen, and L. Zhang, “Data-driven estimation of building energy consumption and GHG emissions using explainable artificial intelligence,” *Energy*, vol. 262, p. 125468, Jan. 2023, doi: 10.1016/J.ENERGY.2022.125468.

[60] N. Abbasabadi, M. Ashayeri, R. Azari, B. Stephens, and M. Heidarinejad, “An integrated data-driven framework for urban energy use modeling (UEUM),” *Appl Energy*, vol. 253, p. 113550, Nov. 2019, doi: 10.1016/J.APENERGY.2019.113550.

[61] R. Olu-Ajayi, H. Alaka, I. Sulaimon, F. Sunmola, and S. Ajayi, “Building energy consumption prediction for residential buildings using deep learning and other machine learning techniques,” *Journal of Building Engineering*, vol. 45, p. 103406, Jan. 2022, doi: 10.1016/J.JOBE.2021.103406.

[62] M. Wurm, A. Droin, T. Stark, C. Geiß, W. Sulzer, and H. Taubenböck, “Deep Learning-Based Generation of Building Stock Data from Remote Sensing for Urban Heat Demand Modeling,” *ISPRS International Journal of Geo-Information 2021, Vol. 10, Page 23*, vol. 10, no. 1, p. 23, Jan. 2021, doi: 10.3390/IJGI10010023.

[63] S. Naji *et al.*, “Estimating building energy consumption using extreme learning machine method,” *Energy*, vol. 97, pp. 506–516, Feb. 2016, doi: 10.1016/J.ENERGY.2015.11.037.

[64] A. P. Melo, R. S. Versage, G. Sawaya, and R. Lamberts, “A novel surrogate model to support building energy labelling system: A new approach to assess cooling energy demand in commercial buildings,” *Energy Build*, vol. 131, pp. 233–247, Nov. 2016, doi: 10.1016/J.ENBUILD.2016.09.033.

[65] F. Ascione, N. Bianco, C. De Stasio, G. M. Mauro, and G. P. Vanoli, “Artificial neural networks to predict energy performance and retrofit scenarios for any member of a building category: A novel approach,” *Energy*, vol. 118, pp. 999–1017, Jan. 2017, doi:

10.1016/J.ENERGY.2016.10.126.

[66] S. M. C. Magalhães, V. M. S. Leal, and I. M. Horta, "Modelling the relationship between heating energy use and indoor temperatures in residential buildings through Artificial Neural Networks considering occupant behavior," *Energy Build*, vol. 151, pp. 332–343, Sep. 2017, doi: 10.1016/J.ENBUILD.2017.06.076.

[67] S. Papadopoulos, E. Azar, W. L. Woon, and C. E. Kontokosta, "Evaluation of tree-based ensemble learning algorithms for building energy performance estimation," *J Build Perform Simul*, vol. 11, no. 3, pp. 322–332, May 2018, doi: 10.1080/19401493.2017.1354919.

[68] M. N. Lopes and R. Lamberts, "Development of a Metamodel to Predict Cooling Energy Consumption of HVAC Systems in Office Buildings in Different Climates," *Sustainability 2018, Vol. 10, Page 4718*, vol. 10, no. 12, p. 4718, Dec. 2018, doi: 10.3390/SU10124718.

[69] A. S. Jihad and M. Tahiri, "Forecasting the heating and cooling load of residential buildings by using a learning algorithm 'gradient descent', Morocco," *Case Studies in Thermal Engineering*, vol. 12, pp. 85–93, Sep. 2018, doi: 10.1016/J.CSITE.2018.03.006.

[70] N. T. Ngo, "Early predicting cooling loads for energy-efficient design in office buildings by machine learning," *Energy Build*, vol. 182, pp. 264–273, Jan. 2019, doi: 10.1016/J.ENBUILD.2018.10.004.

[71] A. D'Amico, G. Ciulla, M. Traverso, V. Lo Brano, and E. Palumbo, "Artificial Neural Networks to assess energy and environmental performance of buildings: An Italian case study," *J Clean Prod*, vol. 239, p. 117993, Dec. 2019, doi: 10.1016/J.JCLEPRO.2019.117993.

[72] J. Vazquez-Canteli, A. Dilsiz Demir, J. Brown, and Z. Nagy, "Deep neural networks as surrogate models for urban energy simulations," *J Phys Conf Ser*, vol. 1343, no. 1, p. 012002, Nov. 2019, doi: 10.1088/1742-6596/1343/1/012002.

[73] G. Ciulla, A. D'Amico, V. Lo Brano, and M. Traverso, "Application of optimized artificial intelligence algorithm to evaluate the heating energy demand of non-residential buildings at European level," *Energy*, vol. 176, pp. 380–391, Jun. 2019, doi: 10.1016/J.ENERGY.2019.03.168.

[74] P. Westermann, M. Welzel, and R. Evins, "Using a deep temporal convolutional network as a building energy surrogate model that spans multiple climate zones," *Appl Energy*, vol.

278, p. 115563, Nov. 2020, doi: 10.1016/J.APENERGY.2020.115563.

[75] X. J. Luo, L. O. Oyedele, A. O. Ajayi, and O. O. Akinade, “Comparative study of machine learning-based multi-objective prediction framework for multiple building energy loads,” *Sustain Cities Soc*, vol. 61, p. 102283, Oct. 2020, doi: 10.1016/J.SCS.2020.102283.

[76] K. W. Mui, L. T. Wong, M. K. Satheesan, and A. Balachandran, “A Hybrid Simulation Model to Predict the Cooling Energy Consumption for Residential Housing in Hong Kong,” *Energies 2021, Vol. 14, Page 4850*, vol. 14, no. 16, p. 4850, Aug. 2021, doi: 10.3390/EN14164850.

[77] E. Elbeltagi and H. Wefki, “Predicting energy consumption for residential buildings using ANN through parametric modeling,” *Energy Reports*, vol. 7, pp. 2534–2545, Nov. 2021, doi: 10.1016/J.EGYR.2021.04.053.

[78] X. Li and R. Yao, “Modelling heating and cooling energy demand for building stock using a hybrid approach,” *Energy Build*, vol. 235, p. 110740, Mar. 2021, doi: 10.1016/J.ENBUILD.2021.110740.

[79] C. Lee, D. E. Jung, D. Lee, K. H. Kim, and S. L. Do, “Prediction Performance Analysis of Artificial Neural Network Model by Input Variable Combination for Residential Heating Loads,” *Energies 2021, Vol. 14, Page 756*, vol. 14, no. 3, p. 756, Feb. 2021, doi: 10.3390/EN14030756.

[80] Y. Liu, H. Chen, L. Zhang, and Z. Feng, “Enhancing building energy efficiency using a random forest model: A hybrid prediction approach,” *Energy Reports*, vol. 7, pp. 5003–5012, Nov. 2021, doi: 10.1016/J.EGYR.2021.07.135.

[81] 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,” *J Build Perform Simul*, vol. 15, no. 4, pp. 553–573, Jul. 2022, doi: 10.1080/19401493.2021.2001573.

[82] J. M. Santos-Herrero, J. M. Lopez-Guede, I. Flores Abascal, and E. Zulueta, “Energy and thermal modelling of an office building to develop an artificial neural networks model,” *Scientific Reports 2022 12:1*, vol. 12, no. 1, pp. 1–13, May 2022, doi: 10.1038/s41598-022-12924-9.

[83] H. Yoshino, T. Hong, and N. Nord, “IEA EBC annex 53: Total energy use in buildings—Analysis and evaluation methods,” *Energy Build*, vol. 152, pp. 124–136, Oct. 2017, doi:

10.1016/J.ENBUILD.2017.07.038.

[84] S. Chen, G. Zhang, X. Xia, Y. Chen, S. Setunge, and L. Shi, “The impacts of occupant behavior on building energy consumption: A review,” *Sustainable Energy Technologies and Assessments*, vol. 45, p. 101212, Jun. 2021, doi: 10.1016/J.SETA.2021.101212.

[85] J. Li, Z. (Jerry) Yu, F. Haghighat, and G. Zhang, “Development and improvement of occupant behavior models towards realistic building performance simulation: A review,” *Sustain Cities Soc*, vol. 50, p. 101685, Oct. 2019, doi: 10.1016/J.SCS.2019.101685.

[86] H. B. Gunay, W. O’Brien, and I. Beausoleil-Morrison, “A critical review of observation studies, modeling, and simulation of adaptive occupant behaviors in offices,” *Build Environ*, vol. 70, pp. 31–47, Dec. 2013, doi: 10.1016/J.BUILDENV.2013.07.020.

[87] G. Happle, J. A. Fonseca, and A. Schlueter, “A review on occupant behavior in urban building energy models,” *Energy Build*, vol. 174, pp. 276–292, Sep. 2018, doi: 10.1016/J.ENBUILD.2018.06.030.

[88] D. Yan *et al.*, “Occupant behavior modeling for building performance simulation: Current state and future challenges,” *Energy Build*, vol. 107, pp. 264–278, Nov. 2015, doi: 10.1016/J.ENBUILD.2015.08.032.

[89] A. Wagner, W. O’brien, and B. Dong, “Exploring Occupant Behavior in Buildings Methods and Challenges”.

[90] W. O’Brien *et al.*, “Introducing IEA EBC annex 79: Key challenges and opportunities in the field of occupant-centric building design and operation,” *Build Environ*, vol. 178, p. 106738, Jul. 2020, doi: 10.1016/J.BUILDENV.2020.106738.

[91] K. Amasyali and N. M. El-Gohary, “A review of data-driven building energy consumption prediction studies,” *Renewable and Sustainable Energy Reviews*, vol. 81, pp. 1192–1205, Jan. 2018, doi: 10.1016/J.RSER.2017.04.095.

[92] T. Hong, D. Yan, S. D’Oca, and C. fei Chen, “Ten questions concerning occupant behavior in buildings: The big picture,” *Build Environ*, vol. 114, pp. 518–530, Mar. 2017, doi: 10.1016/J.BUILDENV.2016.12.006.

[93] Q. Darakdjian, S. Billé, and C. Inard, “Data mining of building performance simulations comprising occupant behaviour modelling,” *https://doi.org/10.1080/17512549.2017.1421099*, vol. 13, no. 2, pp. 157–173, 2018, doi: 10.1080/17512549.2017.1421099.

[94] S. Carlucci, F. Causone, S. Biandrate, M. Ferrando, A. Moazami, and S. Erba, "On the impact of stochastic modeling of occupant behavior on the energy use of office buildings," *Energy Build*, vol. 246, p. 111049, Sep. 2021, doi: 10.1016/J.ENBUILD.2021.111049.

[95] J. Mahecha Zambrano, U. Filippi Oberegger, and G. Salvalai, "Towards integrating occupant behaviour modelling in simulation-aided building design: Reasons, challenges and solutions," *Energy Build*, vol. 253, p. 111498, Dec. 2021, doi: 10.1016/J.ENBUILD.2021.111498.

[96] I. Gaetani, P. J. Hoes, and J. L. M. Hensen, "Occupant behavior in building energy simulation: Towards a fit-for-purpose modeling strategy," *Energy Build*, vol. 121, pp. 188–204, Jun. 2016, doi: 10.1016/J.ENBUILD.2016.03.038.

[97] B. Dong *et al.*, "A guideline to document occupant behavior models for advanced building controls," *Build Environ*, vol. 219, p. 109195, Jul. 2022, doi: 10.1016/J.BUILDENV.2022.109195.

[98] G. Maracchini, E. Di Giuseppe, and M. D'Orazio, "Impact of Occupants' Behavior Uncertainty on Building Energy Consumption Through the Karhunen-Loève Expansion Technique: A Case Study in Italy," *Smart Innovation, Systems and Technologies*, vol. 263, pp. 197–207, 2022, doi: 10.1007/978-981-16-6269-0_17/FIGURES/3.

[99] M. Eguaras-Martínez, M. Vidaurre-Arbizu, and C. Martín-Gómez, "Simulation and evaluation of Building Information Modeling in a real pilot site," *Appl Energy*, vol. 114, pp. 475–484, Feb. 2014, doi: 10.1016/J.APENERGY.2013.09.047.

[100] Z. Li, Y. Jiang, and QP Wei, "'Survey on energy consumption of air conditioning in summer in a residential building in Beijing,'" *J. Heat. Vent. Air Cond*, vol. 37, pp. 46–51, 2007, Accessed: May 17, 2023. [Online]. Available: https://scholar.google.com/scholar?hl=en&as_sdt=0%2C5&q=Z.+Li%2C+Y.+Jiang%2C+Q.+P.+Wei+and+others%2C+%22Survey+on+energy+consumption+of+air+conditioning+in+summer+in+a+residential+building+in+Beijing%2C%22+J.+Heat.+Vent.+Air+Cond%2C+vol.+37%2C+p.+46%E2%80%9351%2C+2007.+&btnG=

[101] O. Ahmed, N. Sezer, M. Ouf, L. (Leon) Wang, and I. G. Hassan, "State-of-the-art review of occupant behavior modeling and implementation in building performance simulation," *Renewable and Sustainable Energy Reviews*, vol. 185, p. 113558, Oct. 2023, doi: 10.1016/J.RSER.2023.113558.

[102] X. Zhou, D. Yan, T. Hong, and X. Ren, “Data analysis and stochastic modeling of lighting energy use in large office buildings in China,” *Energy Build*, vol. 86, pp. 275–287, Jan. 2015, doi: 10.1016/J.ENBUILD.2014.09.071.

[103] R. Melfi, B. Rosenblum, B. Nordman, and K. Christensen, “Measuring building occupancy using existing network infrastructure,” *2011 International Green Computing Conference and Workshops, IGCC 2011*, 2011, doi: 10.1109/IGCC.2011.6008560.

[104] V. Fabi, R. V. Andersen, S. Corgnati, and B. W. Olesen, “Occupants’ window opening behaviour: A literature review of factors influencing occupant behaviour and models,” *Build Environ*, vol. 58, pp. 188–198, Dec. 2012, doi: 10.1016/J.BUILDENV.2012.07.009.

[105] Y. Zhang and P. Barrett, “Factors influencing occupants’ blind-control behaviour in a naturally ventilated office building,” *Build Environ*, vol. 54, pp. 137–147, Aug. 2012, doi: 10.1016/J.BUILDENV.2012.02.016.

[106] J. Rouleau, L. Gosselin, and P. Blanchet, “Robustness of energy consumption and comfort in high-performance residential building with respect to occupant behavior,” *Energy*, vol. 188, p. 115978, Dec. 2019, doi: 10.1016/J.ENERGY.2019.115978.

[107] S. Gilani, W. O’Brien, H. B. Gunay, and J. S. Carrizo, “Use of dynamic occupant behavior models in the building design and code compliance processes,” *Energy Build*, vol. 117, pp. 260–271, Apr. 2016, doi: 10.1016/J.ENBUILD.2015.10.044.

[108] T. Picard, T. Hong, N. Luo, S. H. Lee, and K. Sun, “Robustness of energy performance of Zero-Net-Energy (ZNE) homes,” *Energy Build*, vol. 224, p. 110251, Oct. 2020, doi: 10.1016/J.ENBUILD.2020.110251.

[109] Y. Zhang, X. Bai, F. P. Mills, and J. C. V. Pezzey, “Rethinking the role of occupant behavior in building energy performance: A review,” *Energy Build*, vol. 172, pp. 279–294, Aug. 2018, doi: 10.1016/J.ENBUILD.2018.05.017.

[110] E. Azar *et al.*, “Simulation-aided occupant-centric building design: A critical review of tools, methods, and applications,” *Energy Build*, vol. 224, p. 110292, Oct. 2020, doi: 10.1016/J.ENBUILD.2020.110292.

[111] S. Nejadshamsi, U. Eicker, C. Wang, and J. Bentahar, “Data sources and approaches for building occupancy profiles at the urban scale – A review,” *Build Environ*, vol. 238, p. 110375, Jun. 2023, doi: 10.1016/J.BUILDENV.2023.110375.

[112] F. D. Salim *et al.*, “Modelling urban-scale occupant behaviour, mobility, and energy in

buildings: A survey," *Build Environ*, vol. 183, p. 106964, Oct. 2020, doi: 10.1016/J.BUILDENV.2020.106964.

[113] R. Baetens and D. Saelens, "Modelling uncertainty in district energy simulations by stochastic residential occupant behaviour," *http://dx.doi.org/10.1080/19401493.2015.1070203*, vol. 9, no. 4, pp. 431–447, Jul. 2015, doi: 10.1080/19401493.2015.1070203.

[114] G. Happle, J. A. Fonseca, and A. Schlueter, "Impacts of diversity in commercial building occupancy profiles on district energy demand and supply," *Appl Energy*, vol. 277, p. 115594, Nov. 2020, doi: 10.1016/J.APENERGY.2020.115594.

[115] Y. Wu, J. An, M. Qian, and D. Yan, "Application-driven level-of-detail modeling framework for occupant air-conditioning behavior in district cooling," *Journal of Building Engineering*, vol. 70, p. 106401, Jul. 2023, doi: 10.1016/J.JOBE.2023.106401.

[116] U. Ali, M. H. Shamsi, C. Hoare, E. Mangina, and J. O'Donnell, "A data-driven approach for multi-scale building archetypes development," *Energy Build*, vol. 202, p. 109364, Nov. 2019, doi: 10.1016/J.ENBUILD.2019.109364.

[117] M. H. Kristensen, R. E. Hedegaard, and S. Petersen, "Hierarchical calibration of archetypes for urban building energy modeling," *Energy Build*, vol. 175, pp. 219–234, Sep. 2018, doi: 10.1016/J.ENBUILD.2018.07.030.

[118] "ASHRAE standard. 90.1-2016. User's manual : ANSI/ASHRAE/IES standard 90.1-2016 : energy standard for buildings except low-rise residential buildings." Accessed: Jan. 15, 2023. [Online]. Available: https://www.worldcat.org/title/ashrae-standard-901-2010-users-manual-ansiashraeies-standard-901-2010-energy-standard-for-buildings-except-low-rise-residential-buildings/oclc/896192799

[119] M. D. McKay, R. J. Beckman, and W. J. Conover, "A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code," *Technometrics*, vol. 21, no. 2, p. 239, May 1979, doi: 10.2307/1268522.

[120] J. C. Helton, J. D. Johnson, C. J. Sallaberry, and C. B. Storlie, "Survey of sampling-based methods for uncertainty and sensitivity analysis," *Reliab Eng Syst Saf*, vol. 91, no. 10–11, pp. 1175–1209, Oct. 2006, doi: 10.1016/J.RESS.2005.11.017.

[121] "A Toolkit for Using Whole Building Simulation Program 'EnergyPlus' [R package eplusr version 0.16.2]," Aug. 2023, Accessed: May 07, 2024. [Online]. Available:

https://CRAN.R-project.org/package=eplusr

[122] A. Afshari, C. Nikolopoulou, and M. Martin, “Life-Cycle Analysis of Building Retrofits at the Urban Scale—A Case Study in United Arab Emirates,” *Sustainability 2014, Vol. 6, Pages 453-473*, vol. 6, no. 1, pp. 453–473, Jan. 2014, doi: 10.3390/SU6010453.

[123] N. D. Roman, F. Bre, V. D. Fachinotti, and R. Lamberts, “Application and characterization of metamodels based on artificial neural networks for building performance simulation: A systematic review,” *Energy Build*, vol. 217, p. 109972, Jun. 2020, doi: 10.1016/J.ENBUILD.2020.109972.

[124] J. L. Loeppky, J. Sacks, and W. J. Welch, “Choosing the Sample Size of a Computer Experiment: A Practical Guide,” *Technometrics*, vol. 51, no. 4, pp. 366–376, Nov. 2009, doi: 10.1198/TECH.2009.08040.

[125] C. M. Calama-González, P. Symonds, G. Petrou, R. Suárez, and Á. L. León-Rodríguez, “Bayesian calibration of building energy models for uncertainty analysis through test cells monitoring,” *Appl Energy*, vol. 282, p. 116118, Jan. 2021, doi: 10.1016/J.APENERGY.2020.116118.

[126] “R: Optimal Augmented Latin Hypercube Sample.” Accessed: May 07, 2024. [Online]. Available: https://search.r-project.org/CRAN/refmans/lhs/html/optAugmentLHS.html

[127] “1.17. Neural network models (supervised) — scikit-learn 1.4.2 documentation.” Accessed: May 12, 2024. [Online]. Available: https://scikit-learn.org/stable/modules/neural_networks_supervised.html

[128] G. Panchal, A. Ganatra, … Y. K.-I. J. of, and undefined 2011, “Behaviour analysis of multilayer perceptrons with multiple hidden neurons and hidden layers,” *researchgate.netG Panchal, A Ganatra, YP Kosta, D PanchalInternational Journal of Computer Theory and Engineering, 2011•researchgate.net*, vol. 3, no. 2, pp. 332–337, 2011, doi: 10.7763/IJCTE.2011.V3.328.

[129] S. Haykin, *Neural networks: a comprehensive foundation*. 1998. Accessed: May 19, 2024. [Online]. Available: https://dl.acm.org/doi/abs/10.5555/521706

[130] “Joblib: running Python functions as pipeline jobs — joblib 1.5.dev0 documentation.” Accessed: May 28, 2024. [Online]. Available: https://joblib.readthedocs.io/en/latest/index.html#module-joblib

[131] “xlwings Documentation.” Accessed: May 28, 2024. [Online]. Available:

https://docs.xlwings.org/en/latest/

[132] A. Gastli, Y. Charabi, R. A. Alammari, and A. M. Al-Ali, "Correlation between climate data and maximum electricity demand in Qatar," *2013 7th IEEE GCC Conference and Exhibition, GCC 2013*, pp. 565–570, 2013, doi: 10.1109/IEEEGCC.2013.6705841.

[133] M. Krarti and P. Ihm, "Evaluation of net-zero energy residential buildings in the MENA region," *Sustain Cities Soc*, vol. 22, pp. 116–125, Apr. 2016, doi: 10.1016/J.SCS.2016.02.007.

[134] I. Andric, A. Kamal, and S. G. Al-Ghamdi, "Efficiency of green roofs and green walls as climate change mitigation measures in extremely hot and dry climate: Case study of Qatar," *Energy Reports*, vol. 6, pp. 2476–2489, Nov. 2020, doi: 10.1016/J.EGYR.2020.09.006.

[135] K. Ahmed, L. E. Ortiz, and J. E. Gonzalez, "On the Spatio-Temporal End-User Energy Demands of a Dense Urban Environment," *Journal of Solar Energy Engineering, Transactions of the ASME*, vol. 139, no. 4, Aug. 2017, doi: 10.1115/1.4036545/379727.

[136] L. E. Ortiz, J. E. Gonzalez, E. Gutierrez, and M. Arend, "Forecasting Building Energy Demands with a Coupled Weather-Building Energy Model in a Dense Urban Environment," *Journal of Solar Energy Engineering, Transactions of the ASME*, vol. 139, no. 1, Feb. 2017, doi: 10.1115/1.4034909/377848.

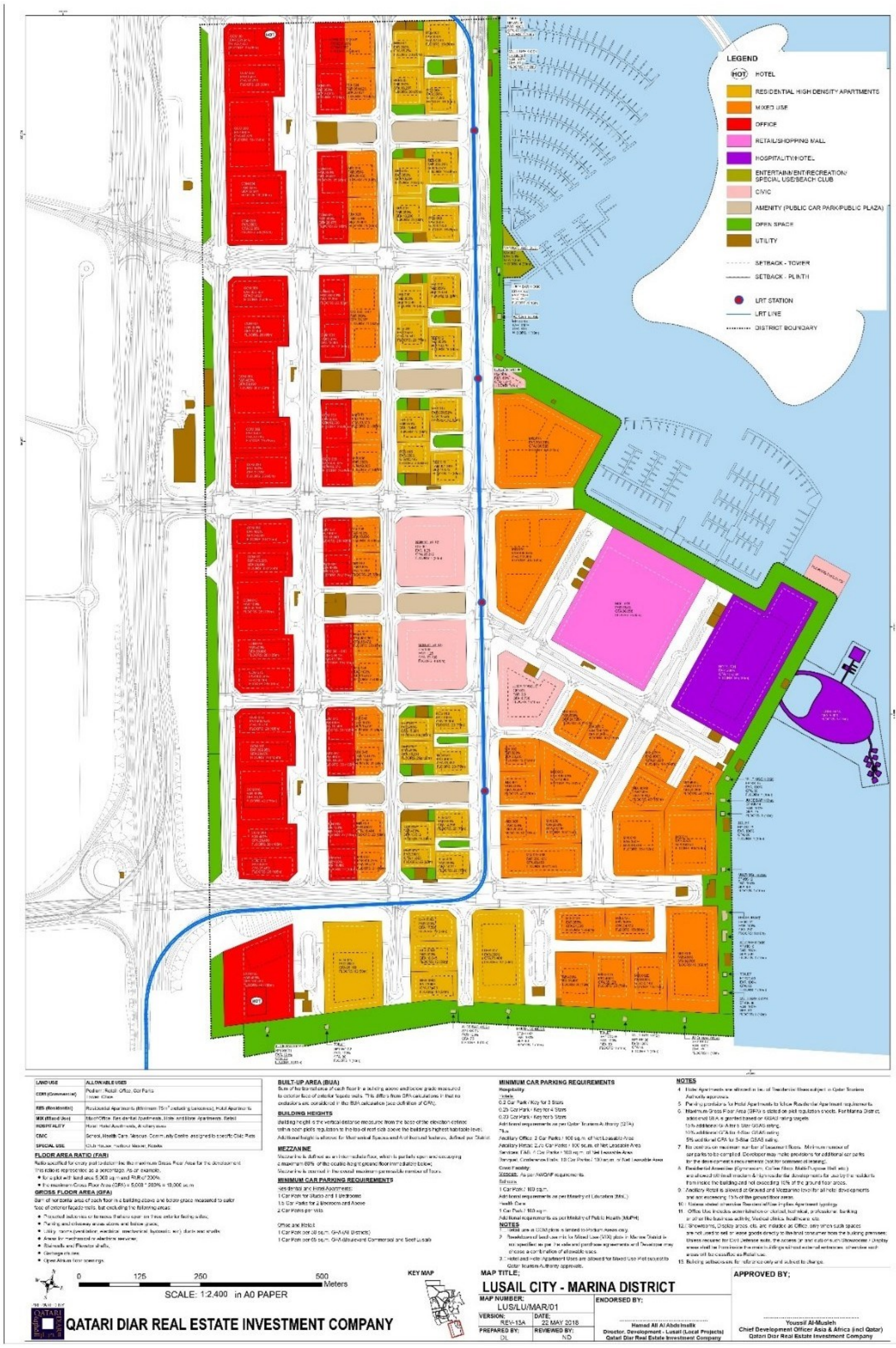
[137] E. Solgi, R. Fayaz, and B. M. Kari, "Cooling load reduction in office buildings of hot-arid climate, combining phase change materials and night purge ventilation," *Renew Energy*, vol. 85, pp. 725–731, Jan. 2016, doi: 10.1016/J.RENENE.2015.07.028.

[138] E. Azar, B. Alaifan, M. Lin, E. Trepci, and M. El Asmar, "Drivers of energy consumption in Kuwaiti buildings: Insights from a hybrid statistical and building performance simulation approach," *Energy Policy*, vol. 150, p. 112154, Mar. 2021, doi: 10.1016/J.ENPOL.2021.112154.

[139] N. Alkaabi, C. S. Cho, A. Mayyas, and E. Azar, "A data-driven modeling and analysis approach to test the resilience of green buildings to uncertainty in operation patterns," *Energy Sci Eng*, vol. 8, no. 12, pp. 4250–4269, Dec. 2020, doi: 10.1002/ESE3.808.

7. Appendix

Appendix 1: Marina District development plan



Appendix 2: Geometric characteristics for the buildings of the Marina District

Building ID	Building type	Floor number	Aspect ratio	GFA (m^2)
1	Residential	20	1	18,234
2	Residential	16	1.2	17,343
3	Residential	20	1	18,207
4	Residential	16	1.17	17,693
5	Residential	20	1	18,204
6	Residential	16	1.2	17,679
7	Residential	21	1	18,200
8	Residential	16	1.2	17,658
9	Residential	20	1	16,590
10	Residential	16	1	16,664
11	Residential	20	1	16,541
12	Residential	16	1	16,520
13	Residential	23	1	15,446
14	Residential	17	1	16,629
15	Residential	23	1	15,446
16	Residential	17	1	16,622
17	Residential	29	1.44	15,201
18	Residential	20	1	16,304
19	Residential	29	1.44	15,201
20	Residential	20	1	16,352
21	Residential	25	1.44	13,512
22	Residential	20	1	14,298
23	Residential	25	1.44	11,823
24	Residential	20	1	14,217
25	Residential	12	1.13	85,158
26	Residential	12	1.12	17,580
27	Residential	12	1.65	18,045
28	Residential	12	1.12	17,403
29	Residential	12	1.12	71,404
1	Commercial	25	1	32,753
2	Commercial	25	1	32,753
3	Commercial	36	1	60,876
4	Commercial	28	1	32,981
5	Commercial	28	1	32,956
6	Commercial	25	1	31,462
7	Commercial	25	1.1	31,469
8	Commercial	36	1.27	58,160
9	Commercial	25	1.1	31,465
10	Commercial	25	1.1	31,367
11	Commercial	33	1.65	29,500
12	Commercial	33	1.65	29,496
13	Commercial	28	1.54	50,164
14	Commercial	28	1.65	29,488
15	Commercial	28	1.65	29,488
16	Commercial	28	1.65	29,476
17	Commercial	28	1.65	29,476

18	Commercial	42	1.54	50,152
19	Commercial	28	1.65	29,480
20	Commercial	28	2.28	29,489
21	Commercial	22	1.3	41,587
22	Commercial	22	1.2	20,146
23	Commercial	22	1.3	20,472
24	Commercial	22	1.26	20,475
25	Commercial	22	1.14	19,635
26	Commercial	22	1	18,995
27	Commercial	23	1.12	19,336
28	Commercial	23	1.4	19,376
29	Commercial	28	1.1	16,840
30	Commercial	28	1	16,120
31	Commercial	39	1	41,180
32	Commercial	39	1	41,180
33	Commercial	36	1	16,568
34	Commercial	36	1	16,592
35	Commercial	27	1	16,472
36	Commercial	27	1	16,460
37	Commercial	48	2.2	100,480
1	Mixed-use	22	1.1	41,587
2	Mixed-use	21	1	20,157
3	Mixed-use	21	1.1	20,430
4	Mixed-use	21	1.1	20,461
5	Mixed-use	21	1	38,367
6	Mixed-use	21	1.1	38,367
7	Mixed-use	21	1	19,373
8	Mixed-use	21	1	19,005
9	Mixed-use	25	1.1	16,868
10	Mixed-use	25	1	15,812
11	Mixed-use	25	1.1	16,472
12	Mixed-use	25	1	16,472
13	Mixed-use	22	1.1	16,204
14	Mixed-use	22	1.1	16,592
15	Mixed-use	22	1	16,444
16	Mixed-use	21	1.1	16,472
17	Mixed-use	60	1.1	166,565
18	Mixed-use	60	1.13	166,565
19	Mixed-use	60	1.2	166,565
20	Mixed-use	40	1.1	112,200
21	Mixed-use	60	1	166,565
22	Mixed-use	60	1.3	166,565
23	Mixed-use	30	1.1	24,720
24	Mixed-use	30	1.1	24,800
25	Mixed-use	40	1	28,285
26	Mixed-use	35	1	43,980
27	Mixed-use	40	1	28,475
28	Mixed-use	30	1	20,230
29	Mixed-use	30	1.1	20,230

30	Mixed-use	40	1.14	21,228
31	Mixed-use	40	1.35	28,844
32	Mixed-use	40	2	32,696
33	Mixed-use	35	1.35	49,728
34	Mixed-use	40	1.35	49,658
35	Mixed-use	15	1.7	24,335
36	Mixed-use	15	1.4	24,772
37	Mixed-use	20	1	35,400
38	Mixed-use	20	1.12	35,128
39	Mixed-use	20	1.22	35,140
40	Mixed-use	40	2.5	92,708
41	Mixed-use	30	3.5	40,460