

A Surrogate Modelling Framework for Time-Resolved Energy Prediction in High-Performance Office
Buildings

by

Ehsan Akbari

A thesis

presented to the University of Waterloo

in fulfillment of the

thesis requirement for the degree of

Doctor of Philosophy

in

Civil Engineering

Waterloo, Ontario, Canada, 2026

© Ehsan Akbari 2026

Examining Committee Membership

The following served on the Examining Committee for this thesis. The decision of the Examining Committee is by majority vote. External Examiner

Chair	Name: Mark Crowley Title: Associate Professor Department of Electrical and Computer Engineering, University of Waterloo
Supervisor	Name: Costa Kapsis Title: Assistant Professor Department of Civil and Environmental Engineering, University of Waterloo
Internal Member	Name: Joyce Kim Title: Assistant Professor Department of Civil and Environmental Engineering, University of Waterloo
Internal Member	Name: Chul Min Yeum Title: Assistant Professor Department of Civil and Environmental Engineering, University of Waterloo
Internal-external Member	Name: Mehrdad Pirnia Title: Professor Department of Management Science and Engineering, University of Waterloo
External Member	Name: Burak Gunay Title: Associate Professor Department of Civil and Environmental Engineering, Carleton University

Author's Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

High-performance buildings are an important component of the transition toward low-carbon and net-zero energy systems. However, evaluating building energy performance across a wide range of design and operational conditions typically requires large numbers of detailed simulations, which can be computationally expensive. This research investigates the energy behavior of high-performance office buildings and develops a surrogate modelling framework capable of predicting time-resolved energy consumption using machine learning techniques.

The study follows a four-phase methodology. In the first phase, operational data from a high-performance office building located in Waterloo, Ontario are analyzed to examine daily energy-use behavior and identify the primary operational drivers of electricity consumption. This analysis provides empirical insight that informs the selection of key parameters used in subsequent modelling stages.

In the second phase, a parametric physics-based building energy model is developed in EnergyPlus to represent office buildings operating under Waterloo's climate conditions. The model incorporates variations in building geometry, envelope properties and configurations, and internal loads to represent a range of plausible design and operational scenarios.

In the third phase, the parametric model is used to generate a synthetic dataset through systematic sampling of the input parameter space. The resulting dataset contains hourly energy simulation outputs across a wide range of building configurations and serves as the training and evaluation data for surrogate models.

In the fourth phase, machine learning algorithms are developed, trained, and evaluated to predict building energy performance, including hourly electricity consumption and annual performance indicators. The trained models are also used to examine the relative influence of key building parameters on predicted energy outcomes.

Three surrogate architectures were evaluated: a fully connected artificial neural network (ANN), a convolutional neural network (CNN), and a hybrid ANN–CNN model. The ANN showed stronger performance for annual energy prediction but limited accuracy in reproducing hourly temporal patterns, while the CNN captured hourly variations more effectively but produced less accurate annual energy estimates. The hybrid ANN–CNN model combined these complementary strengths, achieving comparable hourly prediction accuracy to the CNN and similar annual prediction performance to the ANN. The hybrid model achieved an hourly RMSE of approximately 12.89 kWh and an annual energy prediction RMSE of 1.08 kWh/m²·yr, providing the most balanced overall performance among the evaluated architectures. However, the surrogate models showed a tendency to underestimate peak energy demand, indicating limitations in capturing short-duration peak loads.

Acknowledgements

I would like to express my deepest gratitude to everyone who supported me throughout this journey.

First and foremost, I extend my heartfelt thanks to my family for their unwavering encouragement, patience, and belief in me. Your support has been my foundation. To my wife, whose love, understanding, and constant motivation sustained me through the most challenging moments, thank you. This achievement would not have been possible without you.

I am sincerely grateful to my committee members, Dr. Burak Gunay, Dr. Mehrdad Pirnia, Dr. Chul Min Yeum, and Dr. Joyce Kim, for their valuable time, guidance, and thoughtful feedback. Their expertise strengthened my work and contributed greatly to the completion of this thesis.

I would also like to express my profound appreciation to my supervisor, Dr. Costa Kapsis, for his continuous support, mentorship, and encouragement throughout this entire journey. His insights, patience, and dedication have been instrumental to both my research and my personal development as a scholar.

I am very grateful for the insights and support I received from Dr. Harry Vallianos. It was his knowledge sharing that helped me develop the parametric numerical model.

I also wish to thank Professor David Mather for his support and insights on the case study analysis. Furthermore, I thank Thor Neumann and the Cora Group for sharing performance data and insights on the case study building (Evol1).

To all of you, thank you for helping me reach this milestone.

Table of Contents

List of Figures	xi
List of Tables	xiv
List of Abbreviations	xv
List of Symbols	xvii
Latin Symbols.....	xvii
Greek Symbols	xvii
Subscripts	xviii
1 Introduction.....	1
1.1 Background.....	1
1.1.1 Building Performance Simulation in the AEC Industry.....	2
1.1.2 Limitations of Current BPS Practice.....	3
1.1.3 BPS and Building Decarbonization	4
1.2 Research Gap and Problem Statement.....	4
1.2.1 Need for Scalable and Time-Resolved Modelling Approaches	5
1.2.2 Role of Surrogate Modelling in Building Design	5
1.2.3 Research Positioning and Contribution.....	6
1.3 Research Objectives, Scope and Limitations.....	7
1.3.1 Research question	7
1.3.2 Research Objectives	7
1.3.3 Scope	8
1.3.4 Limitations	8
1.4 Thesis Organization.....	8
2 Literature Review.....	10
2.1 High Performance Buildings	10
2.2 Energy Consumption Modelling Methods in Buildings.....	10
2.2.1 Clustering and Regression Approach.....	13
2.3 Building Energy Simulation as the Basis for Surrogate Modelling.....	13
2.4 Surrogate Modelling.....	14
2.4.1 Introduction to Surrogate Modelling.....	14
2.4.2 Model Development Process.....	14
2.4.3 Sampling Strategies and Dataset Size	15
2.4.4 Energy Output Resolution.....	17

2.4.5	Surrogate Modelling Applications	19
2.4.6	Surrogate Model Types	20
2.4.7	Activation Functions, Optimizers, and Kernel Sizes	23
2.4.8	Advantages of Surrogate Modelling	24
2.4.9	Research Gaps.....	25
3	Methodology	27
3.1	Overview	27
3.2	Phase 1 – Data Gathering and Case Study Analysis.....	29
3.2.1	Case Study Description	29
3.2.2	Data Extraction and Cleaning	29
3.2.3	Data Normalization	30
3.2.4	Outlier Detection and Handling	30
3.2.5	Clustering Analysis of Daily Energy Profiles	31
3.2.6	Model Prediction for Each Cluster.....	31
3.2.7	Clustering of Building Load Profiles	32
3.3	Phase 2 – Development of the Numerical Parametric Model.....	33
3.3.1	Simulation Tool Selection.....	33
3.3.2	Parametric EnergyPlus Model.....	33
3.4	Phase 3 – Synthetic Data Generation	35
3.4.1	Sampling Method and Size	35
3.4.2	Data Collection and Storage.....	36
3.5	Phase 4 – Surrogate Model Development and Evaluation	36
3.5.1	Data Splitting	36
3.5.2	Model Selection	37
3.5.3	Hyperparameter Tuning	37
3.5.4	Model Training	38
3.5.5	Model Evaluation.....	38
3.5.6	Model Selection and Comparison:	39
3.5.7	Parameter Sensitivity and Impact Analysis.....	39
3.6	Summary.....	40
4	Investigation of the Energy Behavior of an Office Case Study	41
4.1	Carbon-Neutral Office Case Study.....	41
4.1.1	Building Description	41

4.1.2	PV System Design	44
4.2	Energy Behavior Analysis and Results.....	46
4.2.1	Clustering of the Total Building Energy Consumption Dataset	46
4.2.2	Feature Analysis and Discussion of Individual Loads	50
4.3	Remarks on the Case Study Analysis	61
5	Parametric Numerical Building Energy Model: Development and Verification	63
5.1	Model Description	63
5.1.1	Building Geometry and Thermal Zoning.....	64
5.1.2	Envelope Characteristics	65
5.1.3	Internal Loads and Occupancy Profiles	65
5.1.4	HVAC System Configuration and Performance	65
5.1.5	Parametric Model Inputs	70
5.1.6	Simple variables	71
5.1.7	Geometrical parameters	71
5.1.8	Parameter Range	71
5.1.9	Weather Input.....	75
5.1.10	Simulation Time Steps	75
5.1.11	Model Outputs.....	75
5.1.12	Simulation Failure Handling	75
5.1.13	Python-Based Automation and Validation.....	75
5.1.14	Computational Environment	76
5.2	Numerical Stability and Indoor Environmental Response	76
5.3	Parametric Sensitivity Analysis of the Numerical Energy Model.....	79
5.3.1	Sensitivity Analysis for the Worst-Case Scenario	80
5.3.2	Hypotheses	81
5.3.3	Hypothesis Verification	83
5.3.4	Sensitivity Analysis for the Office Case Study.....	87
6	Results and Analysis	91
6.1	Introduction	91
6.2	Dataset Generation and Preprocessing	91
6.2.1	Dataset Overview	91
6.2.2	Preprocessing and Feature Scaling.....	92
6.3	Model Selection and Structure	92

6.3.1	Baseline Architectures: FC-ANN and CNN	93
6.3.2	Hyperparameter Optimization Process	94
6.3.3	Model Training and Validation of Baseline Architectures	95
6.3.4	Model Performance on the Held-Out Test Set.....	100
6.4	Hybrid ANN–CNN Model	105
6.4.1	Model Architecture (ANN with Convolutional Output Head).....	105
6.4.2	Hyperparameter Optimization for the Hybrid ANN–CNN Model	107
6.4.3	Training and Validation of the Hybrid Model	108
6.4.4	Model Performance on the Held-Out Test Set.....	111
6.5	Final Standalone Model Saving and Post-Processing	114
6.5.1	Standalone Surrogate Model Deployment	114
6.5.2	Post-Processing and Clustering Analysis of Hourly Predictions	115
6.6	Sensitivity Analysis of Surrogate Model.....	120
6.6.1	Overview and Methodology.....	120
6.6.2	Correlation Analysis (Baseline)	120
6.6.3	One-at-a-Time (OAT) Analysis	121
6.6.4	Morris Elementary Effects (Global Sensitivity Analysis).....	123
6.6.5	Permutation Importance Analysis	125
6.6.6	Integrated Sensitivity Framework for Hourly Outputs	126
7	Summary and Conclusions.....	128
7.1	Purpose and Scope of the Study	128
7.2	Summary of Key Findings.....	128
7.2.1	Case Study and Clustering–Regression Analysis.....	128
7.2.2	Surrogate Modelling Framework	129
7.2.3	Sensitivity Analysis and Physical Verification	130
7.3	Contributions	130
7.4	Implications for Design and Operation.....	131
7.5	Limitations.....	131
7.6	Future Work.....	132
7.7	Conclusions	133
	References	134
	Appendices	143
	Appendix Table List	143

Appendix-A: Schedule Input Values	144
Appendix-B: Python Code Requirements	145
Appendix-C: Python Notebooks and GitHub links	147

List of Figures

Figure 1-1. Existing surrogate modelling process for energy consumption in buildings.	6
Figure 2-1- Data types for energy consumption modelling in the buildings [24].	11
Figure 2-2- Summary of the stages in generating a surrogate model and its applications. Adopted from [56].	15
Figure 3-1- A General Four-Phase Framework for Physics-Informed Surrogate Modelling of Building Energy Performance.	29
Figure 3-2- Total energy consumption before (left) and after (right) outlier detection and handling.	31
Figure 3-3- Schematic flowchart of the case study investigation using clustering regression, adapted from [118].	33
Figure 4-1- Carbon neutral office case study, located in southwest Ontario, Canada. [131].	42
Figure 4-2- Simplified depiction of the building HVAC piping system (courtesy of D. Mather).	43
Figure 4-3- Simplified schematic of the air handling unit (AHU) system for the case study building (courtesy of D. Mather).	44
Figure 4-4- Seven clusters for the normalized daily total energy consumption profiles using k-means clustering method. The orange dotted line represents the average daily profile for each cluster.	48
Figure 4-5- Calendar map for total daily energy profiles	49
Figure 4-6- Percentage of individual loads over the total energy consumption of the Office Case Study.	50
Figure 4-7- Calendar map for total energy consumption in 2024.	52
Figure 4-8- HVAC energy clusters using k-means method.	53
Figure 4-9- Calendar map for HVAC energy consumption profiles.	54
Figure 4-10. Proposed normalized daily HVAC schedule during workdays.	54
Figure 4-11- Plug loads clusters using k-means method	55
Figure 4-12- Calendar map for plug loads profiles.	56
Figure 4-13- Representative schedule for periods of increased occupancy, generated based on plug load cluster-PL1 and cluster-PL3.	56
Figure 4-14- A tree flowchart for total energy consumption clusters as appear in Figure 4-8.	61
Figure 5-1- Schematic model geometry and zoning for smallest and largest office buildings represented by the parametric model.	64
Figure 5-2- Occupancy, Lighting and Equipment schedule for office buildings based on NECB (adopted from [10]).	67
Figure 5-3- HVAC setpoint schedule for parametrical numerical model based on high-performance case study building.	68
Figure 5-4- Schematic of VRF system for zones (separated terminal for each zone)	69

Figure 5-5- Schematic of Energy Recovery Ventilation (ERV) system	70
Figure 5-6- Floor zoning scheme defining the minimum building dimensions and resulting minimum floor area based on the aspect ratio.....	73
Figure 5-7- One representative week of hourly setpoint and operative temperature for one of the east perimeter zones during wintertime (heating season).	77
Figure 5-8- Outdoor temperature and global horizontal irradiance for the representative winter week.....	78
Figure 5-9- One representative week of hourly setpoint and operative temperature for one of the east perimeter zones during summertime (cooling season).....	78
Figure 5-10- Outdoor temperature and global horizontal irradiance for the representative summer week	79
Figure 5-11- EUI for different scenarios, based on the total energy in the worst case.	82
Figure 5-12-TEDI for different scenarios, based on the total energy in the worst case.....	82
Figure 5-13-CEDI for different scenarios, based on the total energy in the worst case	83
Figure 5-14- Sensitivity of EUI to variations in wall and roof thermal resistance, with all other variables fixed at baseline values under worse-case conditions,.....	85
Figure 5-15- Sensitivity of TEDI to variations in wall and roof thermal resistance, with all other variables fixed at baseline values under worse-case conditions.....	86
Figure 5-16- Sensitivity of CEDI to variations in wall and roof thermal resistance, with all other variables fixed at baseline values under worse-case conditions.....	86
Figure 5-17- Design parameter sensitivity analysis on EUI	89
Figure 5-18- Design parameter sensitivity analysis on TEDI	90
Figure 5-19- Design parameter sensitivity analysis on CEDI.....	90
Figure 6-1- Total validation losses for fully connected ANN model during training process for all folds	96
Figure 6-2- Total validation losses for CNN model during training process for all folds	96
Figure 6-3- Individual validation losses and loss ratio for fully connected ANN model.	98
Figure 6-4- Individual validation losses and loss ratio for CNN model	99
Figure 6-5- Numerical versus Surrogate for fully connected (FC) ANN model evaluation on test dataset	102
Figure 6-6- Numerical versus Surrogate value for CNN model evaluation on test dataset	104
Figure 6-7- Schematic of hybrid model architecture	106
Figure 6-8- Total validation losses for each fold during training process for Hybrid ANN-CNN model	109
Figure 6-9- Individual validation losses and loss ratio for hybrid ANN-CNN model	110
Figure 6-10- Numerical versus Surrogate for hybrid ANN-CNN model evaluation on test dataset	113

Figure 6-11- Representative daily load profiles and calendar maps from K-means clustering of the hybrid ANN–CNN model’s hourly predictions for a randomly selected input configuration, showing results for the elbow, Calinski–Harabasz, and fixed $k = 5$ configurations.....	118
Figure 6-12- Representative daily load profiles and calendar maps from K-means clustering of the hybrid ANN–CNN model’s hourly predictions for a closest input values to the case study building, showing results for the elbow, Calinski–Harabasz, and fixed $k = 5$ configurations.....	120
Figure 6-13- Correlation coefficients for annual energy metrics and total daily energy	121
Figure 6-14- OAT results for hourly energy outputs and annual energy performance indicators	123
Figure 6-15- Morris sensitivity analysis results for hourly energy outputs	124
Figure 6-16- Permutation importance ranking for the model’s hourly output predictions	126

List of Tables

Table 2-1-EUI for different type of buildings in different climate zones in Canada, adopted from [23]...	10
Table 2-2-Advantage and disadvantages of the building modelling approaches [24], [26].....	12
Table 2-3- Overview of the benefits associated with using surrogate models in the building design [56].	25
Table 3-1- Hyperparameters and their range for FC-ANN.....	37
Table 3-2- Hyperparameters and their range for CNN	38
Table 3-3- Hyperparameters and their range for Hybrid ANN-CNN model.....	38
Table 3-4- Overview of Sensitivity and Importance Analysis Methods.....	40
Table 4-1- Summary of site and building characteristics.....	45
Table 4-2- Plug loads characteristics and its relation to occupancy patterns.....	57
Table 4-3- Prediction results for all the six clusters, with weather parameters as input(s).....	58
Table 4-4- Input parameters and fitting results for polynomial regression and MLP on the total energy consumption of the building.	59
Table 5-1- Input parameters, range, and type of parameters.....	74
Table 5-2- Input parameters base scenario value and their range.....	80
Table 6-1- Optimized hyperparameter values, using Optuna for FC-ANN.....	94
Table 6-2 Optimized hyperparameter values, using Optuna for CNN.....	95
Table 6-3- ANN Ensemble model summary.....	100
Table 6-4- CNN Ensemble model summary.....	100
Table 6-5- Evaluation metrics for fully connected ANN performance on the unseen test dataset	101
Table 6-6- Evaluation metrics for CNN model performance on the unseen test dataset	103
Table 6-7- Optimized hyperparameter values, using Optuna for Hybrid model	107
Table 6-8- Ensemble validation losses for hybrid ANN-CNN model.	111
Table 6-9- Evaluation metrics for Hybrid ANN-CNN performance on the unseen test dataset.....	112
Table 6-10- A set of random input parameters to perform clustering analysis.....	117
Table 6-11- Annual energy indices for the random office building.....	118
Table 6-12- Surrogate input values to the case study building	119
Table 6-13- Sensitivity ranking summary with the normalized value for all methods and normalized composite score.....	127

List of Abbreviations

Abbreviation	Description
AHU	Air Handling Unit
ANN	Artificial Neural Network
ASHRAE	American Society of Heating, Refrigerating and Air-Conditioning Engineers
BC	Branch Circuit
BMS	Building Management System
BPS	Building Performance Simulation
CAGBC	Canada Green Building Council
CDD	Cooling Degree Days ($^{\circ}\text{C}\cdot\text{day}$)
CEDI	Cooling Energy Demand Intensity ($\text{kWh}/\text{m}^2\cdot\text{year}$)
(FC)-CNN	(Fully connected) Convolutional Neural Network
CO ₂	Carbon Dioxide
COP	Coefficient of Performance
CU	Condensing Unit
ECM	Energy Conservation Measure
EPD	Equipment Power Density (W/m^2)
ERV	Energy Recovery Ventilation
EUI	Energy Use Intensity ($\text{kWh}/\text{m}^2\cdot\text{year}$)
FCU	Fan Coil Unit
GHG	Greenhouse gas emissions (kgCO_2e)
HDD	Heating Degree Days ($^{\circ}\text{C}\cdot\text{day}$)
HVAC	Heating, Ventilation and Air Conditioning
IQR	Interquartile range method
LCA	Life Cycle Assessment
LHS	Latin Hypercube Sampling
LPD	Lighting Power Density (W/m^2)
LSTM	Long Short-Term Memory
MAE	Mean Absolute Error (kWh/m^2)
MAPE	Mean Absolute Percentage Error
ML	Machine Learning
MLP	Multilayer Perceptron
NECB	National Energy Code of Canada for Buildings
NZEB	Net-Zero Energy Building
ReLU	Rectified Linear Unit

RF	Random Forrest
RMSE	Relative Mean Squared Error (kWh/m ²)
SHGC	Solar Heat Gain Coefficient
TCN	Temporal Convolutional Network
TEDI	Thermal Energy Demand Intensity
TMY	Typical Meteorological Weather
VRF	Variable Refrigerant Flow
WWR	Window-to-Wall Ratio
XGBoost	Extreme Gradient Boosting
ZCB	Zero Carbon Building

List of Symbols

Latin Symbols

Symbol	Description	Units
A	Surface area	m ²
C	Energy consumption	kWh
E	Energy consumption	kWh
L	Length	m
N_of_Fl	Number of middle floors	(1-10)
O	Normalized occupancy Schedule	(0-1)
P	Normalized PV generation	(0-1)
RSI	Thermal resistance	m ² ·K/W
R ²	Coefficient of determination	–
S	Normalized HVAC set point schedule	(0-1)
T	Temperature	°C
t	Time step	h
U	Thermal transmittance	W/m ² ·K
W	Width	m
X	Datapoint	–

Greek Symbols

Symbol	Description	Units
α	Hourly-to-annual loss weight	–
μ	Mean value	–
σ	Standard deviation	–
Σ	Summation operator	–

Subscripts

Subscript	Description
a	polynomial regression coefficient
C	cooling energy
EV	Electric vehicle
F	floor
H	Heating energy
i	timestep
in	Indoor
L	Lighting energy
max	Maximum
min	Minimum
$norm$	Normalized
out	Outdoor
$peak$	Hourly peak electricity consumption (kWh)
PL	Plug load
PV	Photovoltaic system
r or $roof$	roof
sim	Simulation result
sur	Surrogate model prediction
$test$	Testing dataset
$train$	Training dataset
w or win	window
$zone$	Thermal zone

Chapter 1

1 Introduction

1.1 Background

Climate change, driven by anthropogenic greenhouse gas (GHG) emissions, presents one of the most pressing global challenges. The use of fossil fuels for energy in buildings is a major source of these emissions. Buildings consume energy for heating, cooling, lighting, and other services, and much of this energy comes from fossil fuels, whether directly consumed onsite or as primary energy at the utility level. The resulting emissions contribute significantly to global warming, leading to extreme (in frequency and magnitude) weather events and adverse impacts on ecosystems and human health. As a result, the urgent need to address climate change has driven global efforts to reduce GHG emissions. Globally, buildings are significant contributors, accounting for approximately 36% of global energy-related CO₂ emissions [1]. Urbanization trends are exacerbating the problem. As the global population increasingly migrates to urban areas, the demand for energy in buildings grows. This trend, coupled with rising living standards, intensifies energy consumption and associated emissions. Without intervention, this could lead to an exponential increase in GHG emissions, making it even harder to meet climate targets set by international agreements such as the Paris Accord [2].

Emissions in the Canadian building sector are lower than the world average due to cleaner primary energy sources (i.e., hydro and nuclear), with buildings being responsible for 27%, two thirds of which is related to building operations and one third to the embodied carbon of the material used in the buildings [3]. The following section explores why electrification and decarbonization of buildings are necessary and how net-zero energy buildings (NZEB) and zero-carbon buildings (ZCB) can play a pivotal role in climate mitigation.

In general, electrification refers to shifting energy consumption from fossil fuels (e.g., natural gas for heating and hot water) to electricity, ideally sourced from renewable resources. Decarbonization involves reducing carbon emissions through various means, including energy efficiency, renewable energy adoption, and innovative (passive and active) design strategies. NZEB and ZCB (referred to hereafter as “high-performance buildings”) are designed to minimize energy consumption and offset any remaining energy use with renewable energy generation, resulting in a net zero energy footprint. These buildings not only contribute to the reduction of GHG emissions but also set a precedent for sustainable building practices that can offer enhanced indoor environmental conditions and occupant comfort [4], [5].

These buildings often operate under stringent performance targets such as Energy Use Intensity (EUI), Thermal Energy Demand Intensity (TEDI), and carbon intensity benchmarks. Unlike conventional buildings, their energy behavior is not dominated solely by weather conditions. Instead, operational strategies, internal loads, control schedules, and system interactions play increasingly significant roles in shaping energy demand profiles.

Accurate modelling of such buildings is essential for:

- Early-stage design optimization
- Evaluation of Energy Conservation Measures (ECMs)
- Retrofit analysis
- Operational tuning and performance verification
- Grid-interactive building strategies

Building Performance Simulation (BPS) tools are widely used to support these tasks. These physics-based numerical models provide detailed representations of heat transfer, HVAC systems, and control strategies. However, despite their physical rigor, they exhibit important limitations when used for large-scale scenario exploration or time-resolved performance prediction.

1.1.1 Building Performance Simulation in the AEC Industry

Building performance simulation (BPS) is a fundamental tool in the architecture, engineering, and construction (AEC) industry, supporting the evaluation of building energy use, indoor environmental quality, and environmental impacts across the building life cycle. It plays a critical role in both design and retrofit decision-making by enabling the assessment of alternative design configurations under varying boundary conditions.

During early-stage design, BPS is particularly valuable for comparing building form, envelope characteristics, and system configurations when design information is limited but decisions have a disproportionately large impact on long-term performance. However, at this stage, its application is often constrained by rapidly evolving design parameters, high input uncertainty, and the time-intensive nature of detailed model development [6], [7].

At later design stages, detailed simulation tools such as EnergyPlus [8] are widely used to demonstrate compliance with building energy codes and standards, including ASHRAE Standard 90.1[9] and the National Energy Code of Canada for Buildings(NECB) [10]. While compliance demonstration remains a

common application, BPS is equally essential for multi-objective design optimization, life-cycle assessment, and performance benchmarking. These applications become particularly important when design goals extend beyond minimum code requirements toward net-zero energy and low-carbon performance targets, where regulatory thresholds represent a starting point rather than an endpoint.

Beyond design applications, BPS is increasingly used in retrofit analysis, operational optimization, and performance monitoring. In such cases, simulation models may be calibrated or combined with reduced-order approaches using data from building management systems (BMS) to better represent actual building behavior and reduce the gap between simulated and measured performance [7], [11]

These constraints are particularly consequential for high-performance buildings, where rigorous scenario exploration across a wide design space is often essential to meeting demanding energy and carbon targets.

1.1.2 Limitations of Current BPS Practice

Despite the capabilities outlined above, the practical utility of BPS is curtailed by practical barriers that restrict its application beyond compliance-driven workflows. The development of detailed BPS models requires extensive specification of building geometry, thermal zoning, material properties, internal loads, HVAC systems, and control strategies. This process is inherently labor-intensive and is further complicated by uncertainty in input parameters and frequent design changes, particularly in early project stages [6], [7]. As a result, applying BPS to large-scale parametric or iterative analyses remains challenging in practice.

Parametric studies, optimization, and uncertainty quantification using BPS often require thousands of simulation runs, leading to substantial computational cost. To address this, previous research has explored acceleration strategies such as surrogate (meta-) models, parallel computing, and reduced-order modelling techniques. Nevertheless, these approaches are not yet consistently integrated into standard industry workflows [11], [12].

Consequently, BPS in practice is often applied in a compliance-driven manner, with emphasis placed on aggregated performance indicators such as annual energy use intensity (EUI). This focus can obscure important temporal dynamics, including peak demand, load variability, and interactions between building systems and the occupants [11].

Furthermore, conventional BPS workflows are not well suited for real-time or near-real-time applications, such as operational decision support, demand response, or grid-interactive building controls. These applications require rapid or continuous evaluation of building performance, which is not feasible with traditional simulation approaches [12], [13]. These limitations highlight the need for complementary

modelling approaches that retain physical and performance fidelity while enabling scalable and computationally efficient analysis.

1.1.3 BPS and Building Decarbonization

BPS provides a critical framework for evaluating decarbonization strategies by enabling detailed assessment of energy conservation measures, envelope improvements, HVAC system configurations, and renewable energy integration across the building life cycle. It supports the analysis of trade-offs between competing design options, such as envelope performance versus system efficiency, and enables scenario-based evaluation under future climate and grid conditions [13], [14].

As buildings become more energy-efficient, their performance becomes increasingly sensitive to operational factors, including occupant behavior, control strategies, and system interactions. In addition, the importance of peak demand, load flexibility, and grid interaction is growing in the context of grid-interactive and net-zero energy buildings. These emerging challenges require modelling approaches capable of capturing time-resolved dynamics and complex system interactions while remaining computationally tractable for large-scale analysis [14], [15]. Surrogate modelling, in which machine learning models are trained on high-fidelity simulation outputs (also known as “synthetic data”) has emerged as a promising approach to address precisely these demands and forms the central motivation of the present research.

1.2 Research Gap and Problem Statement

Detailed physics-based BPS tools such as EnergyPlus [8] and TRNSYS [17] are widely used to evaluate design options and operational strategies in high-performance buildings. While these models can represent building physics and system interactions at high resolution, their use in practice is constrained by two structural barriers.

First, developing a credible numerical model is itself a major project cost. A detailed simulation model requires extensive effort to define geometry, zoning, constructions, internal loads, HVAC system configurations, control logic, and schedules. When measured data are available, additional time is required for calibration and verification to reduce the gap between predicted and observed performance. This process demands specialized expertise and can take substantial time, making it impractical to produce high-quality models at scale across many projects. In industry settings, detailed models are frequently developed for a single building to answer a specific question (e.g., compliance, design iteration, retrofit feasibility) and are then rarely reused due to differences in building configuration, incomplete documentation, and calibration requirements.

Second, even after a numerical model is developed, running it repeatedly across a high-dimensional design space is computationally expensive. This limitation becomes critical when the goal is not a single estimate, but broad scenario exploration such as ECM screening, optimization, sensitivity analysis, or evaluation of time-resolved operational dynamics (hourly and daily load profiles). In these cases, the number of simulations required can grow rapidly, making traditional workflows slow and resource intensive.

Surrogate modelling addresses both barriers. A surrogate model is trained on a structured dataset of simulation inputs and outputs and can then reproduce the numerical model's behavior at a fraction of the computational cost.

1.2.1 Need for Scalable and Time-Resolved Modelling Approaches

Purely data-driven approaches can provide fast predictions but may lack physical interpretability, robustness under extrapolation, and reliability when limited measured data are available [6], [17]. This has led to increasing interest in surrogate modelling approaches, in which machine learning or reduced-order models are trained on high-fidelity simulation outputs to approximate building performance at significantly lower computational cost.

Surrogate models have been successfully applied to accelerate design optimization, retrofit analysis, and multi-objective trade-off studies, often achieving orders-of-magnitude reductions in computation time (i.e., from days and several hours to seconds) while maintaining high predictive accuracy for key performance indicators [6], [12], [18], [19]. However, most existing applications focus on aggregated outputs, such as annual energy use or peak demand. Only a limited number of studies address high-dimensional, time-resolved outputs, such as hourly energy profiles, which are essential for operational analysis and grid interaction.

This thesis directly addresses this gap by developing and evaluating a surrogate modelling framework targeting hourly whole-building energy profiles rather than aggregated indicators.

1.2.2 Role of Surrogate Modelling in Building Design

Surrogate models provide a computationally efficient means of approximating high-fidelity building performance simulations, enabling rapid evaluation of design and operational scenarios without repeatedly executing time-intensive numerical models. Surrogate models refer to simplified representations of more complex building models, allowing the rapid evaluation of numerous design scenarios without the need for extensive physical or numerical analysis. These models are typically

utilizing machine learning algorithms trained on synthetic data. The synthetic data is generated using a parametric numerical model. An infographic for creating a surrogate model for building energy consumption is presented in the Figure 1-1. By using surrogate models, researchers can explore the effects of different design parameters and systems on building performance, identifying the most effective combinations for achieving net-zero energy targets.

Beyond rapid prediction, surrogate models support design and operational optimization by enabling extensive parametric studies and sensitivity analyses (e.g., under future climate conditions) across high-dimensional input spaces. This helps in understanding the relative importance and interactions of various factors influencing building performance. Through this process, key design parameters that have the most significant impact on achieving energy (and peak demand power) performance targets can be identified and prioritized. Overall, the use of surrogate building models can enhance the ability to study high-performance building (i.e., NZEB and ZCB) design and operation by providing a powerful tool for design experimentation and optimization, ultimately leading to reliable and effective building performance outcomes.

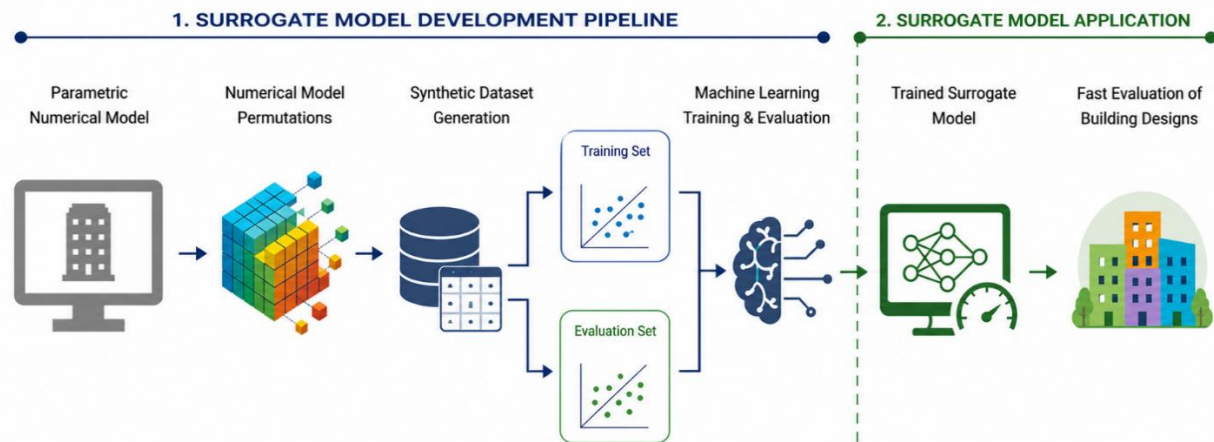


Figure 1-1. Existing surrogate modelling process for energy consumption in buildings.

1.2.3 Research Positioning and Contribution

The absence of time-resolved surrogate modelling approaches is particularly consequential for high-performance buildings, where operational strategies, occupant behavior, and interactions with the power grid play a dominant role in determining energy performance. Capturing these dynamics requires models that can represent both the magnitude and temporal structure of energy use.

This thesis addresses this gap by developing a surrogate modelling framework for predicting hourly energy consumption profiles of high-performance office buildings. Unlike conventional approaches that

focus on scalar outputs, the proposed framework targets high-dimensional temporal outputs, enabling detailed representation of load profiles over time.

This thesis advances the case that surrogate models can reliably and rapidly predict time-resolved, hourly whole-building energy profiles for high-performance office buildings across a wide range of design conditions, at a fraction of the computational cost. The resulting framework enables rapid scenario exploration at a temporal resolution relevant to demand response, grid interaction, and operational decision-making, extending the practical utility of building performance simulation beyond compliance-driven applications.

1.3 Research Objectives, Scope and Limitations

1.3.1 Research question

This thesis aims to answer two primary research questions:

- 1) *What are the main design and operational parameters that significantly impact the energy consumption of a high-performance office building?*
- 2) *Can a surrogate model trained on a parametric simulation dataset accurately predict time-resolved whole-building electricity demand profiles (hourly/daily load shapes) from a defined set of design and operational input parameters, rather than only predicting aggregated indicators such as annual EUI or peak demand?*

1.3.2 Research Objectives

This thesis aims to develop and validate a structured surrogate modelling framework capable of predicting time-resolved energy demand in high-performance office buildings. The work is organized around following defined objectives:

Objective 1: Empirical Characterization of Time-Resolved Energy Behavior: To analyze measured energy data from a high-performance office building and identify dominant daily load patterns and operational drivers using clustering and feature-based interpretation. The empirical characterization of load patterns and operational drivers established in this objective directly defines the input parameter space and output structure of the surrogate model developed in Objective 2

Objective 2: Surrogate Modelling of High-Dimensional Temporal Outputs: Building on the empirical parameter selection and temporal structure identified in Objective 1, this objective aims to develop, train, and evaluate machine learning surrogate models capable of predicting hourly whole-building energy consumption from low-dimensional parametric inputs, and to assess parameter influence using multiple sensitivity analysis methods. This objective constitutes the central research contribution of the thesis.

1.3.3 Scope

This thesis develops and evaluates a surrogate modelling framework for predicting time-resolved operational energy demand in high-performance office buildings located in the Waterloo region (Southern Ontario, Canada). The scope is defined by the following boundaries:

Geographic and Climatic Context: All numerical simulations and surrogate training data are generated using weather conditions representative of the Waterloo region (ASHRAE climate zone 6A). Therefore, model behavior reflects cold-climate operational dynamics typical of Southern Ontario.

Primary Output Variable: The surrogate predicts hourly whole-building electricity **consumption**, from which daily energy profiles and selected annual indicators may be derived. The core contribution lies in high-dimensional temporal output prediction rather than solely annual aggregated metrics.

Intended Application Domain: The framework is intended for:

- rapid parametric scenario exploration,
- preliminary design evaluation,
- ECM screening,
- sensitivity and parameter-prioritization analysis,
- investigation of load-shape behavior relevant to peak demand and operational studies.

The surrogate is not intended to replace detailed building performance simulation for final compliance documentation or regulatory approval.

1.3.4 Limitations

This research is bounded by deliberate scope decisions and two structural methodological constraints. The surrogate framework is developed and evaluated for a single climate zone and building typology; the operational schedule structure is held fixed rather than parameterized; and building geometry is restricted to rectangular footprints with simplified envelope assemblies. Surrogate outputs are scoped to operational energy, excluding cost and emissions. These boundaries are practical in nature, the underlying methodology is transferable, and are discussed in full technical detail in Section 7.5.

1.4 Thesis Organization

This thesis is organized into seven chapters, progressing from theoretical background to methodological development, empirical analysis, surrogate modelling, and final conclusions.

Chapter 2 reviews building energy modelling methods, surrogate modelling, and sensitivity analysis, with emphasis on time-resolved prediction.

Chapter 3 presents the four-phase research methodology: case study analysis, parametric simulation model development, synthetic dataset generation, and surrogate model training/evaluation.

Chapter 4 identifies and interprets recurring daily energy patterns using clustering and feature-based analysis in a high-performance Case Study.

Chapter 5 assesses the numerical model's response to parameter variation and establishes a basis for synthetic dataset generation.

Chapter 6 develops and validates ANN/CNN/hybrid surrogate model for hourly profiles and quantifies parameter influence via sensitivity analyses.

And finally, **Chapter 7** summarizes findings, contributions, limitations, and future work.

Chapter 2

2 Literature Review

This chapter reviews building energy use and the modelling approaches used to predict it, covering physics-based simulation and data-driven methods, discusses key sources of modelling. The review identifies gaps in the current literature, particularly in high-resolution, time-resolved prediction, and establishes the methodological context for the surrogate framework developed in this thesis.

2.1 High Performance Buildings

NZEB and ZCB definitions vary across jurisdictions and research contexts, differing in how they treat balance boundaries, energy carriers, renewable source location, and verification methods. [20] [21] [22].

For this study, a *Net-Zero Energy Building (NZEB)* is defined as a highly efficient, combustion-free, electrified building which produces as much renewable energy on-site as it consumes in a typical year, based on the end-use energy meter, as expressed by equation 2-1.

In Canada, the Canada Green Building Council's Zero Carbon Building standard sets performance targets for thermal energy demand intensity (TEDI) and carbon emission intensity by climate zone and building type [Table 2-1]. For Climate Zone 6 office buildings, the TEDI target is 100 kWh/m²·yr. These targets inform the parameter ranges and output benchmarks used in the surrogate modelling framework developed in this thesis.

$$\sum_i E_{PV_i} - \sum_i (E_{H_i} + E_{C_i} + E_{L_i} + E_{HW_i} + E_{PL_i} + E_{EV_i}) = E_{PV} - E_{LOADS} \geq 0 \quad Eq. 2-1$$

Table 2-1-EUI for different type of buildings in different climate zones in Canada, adopted from [23].

Climate zones	Building type			
	Office	Retail	Multi-unit residential	Hotel/Motel
4	100	90	100	120
5	100	90	110	130
6	100	95	110	130
7 and 8	$0.0078 \cdot HDD^{(1)} + 78$	$0.0068 \cdot HDD + 67$	$0.0116 \cdot HDD + 66$	$0.0091 \cdot HDD + 92$

(1) HDD refers to Heating Degree Days

2.2 Energy Consumption Modelling Methods in Buildings

Understanding the distinction between physics-based simulation and data-driven approaches is fundamental to surrogate modelling research. Physics-based simulation (primarily EnergyPlus) provides

training data for surrogate models, while data-driven methods (ML, DL) constitute the surrogate model architectures themselves. This section establishes the conceptual foundation for the surrogate modelling methodology developed in this thesis.

Coakley et al. [24] presented two different methods for categorizing models of energy consumption:

1. Based on the purpose of the model, it can be categorized as diagnostic or prognostic. Diagnostic models are used to identify the nature or cause of some phenomenon, while prognostic models are used to predict the behavior of a system, given a set of well-defined laws governing that system.

Based on the modelling approach, the model can be categorized as Law-Driven or Data-Driven. Law-Driven (or forward) models apply a given set of physical laws (e.g., gravity, heat/mass transfer etc.) which govern a system, in order to predict its behavior given system properties and conditions. Data-Driven (or inverse) models work on the opposite approach, using system behavior as a predictor for system properties. Therefore, data-driven models can be used to describe a system with a minimal set of adjustable inputs.

Building performance simulation (BPS), which are mostly numerical models, are commonly classified as prognostic law-driven models. Some of the uses of BPS include investigating if conservation measures can be cost effective in the design stage or evaluating the performance of optimization measures in the operation stage of buildings. Since there are a lot of independent parameters which affect energy consumption in a building, it is not a simple job to achieve proper accuracy in energy modelling of real-world buildings [24].

Building energy models draw on three categories of input data: condition data (weather and occupancy), building characteristics (geometry, envelope, systems), and observable operational outputs [25]. These types of data are shown in the Figure 2-1.

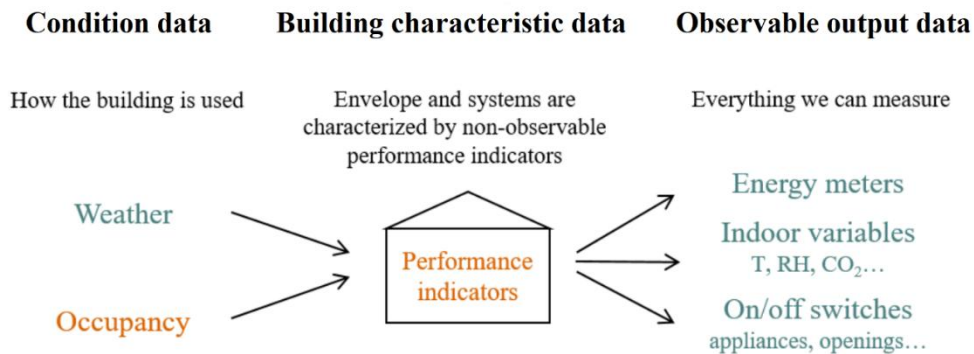


Figure 2-1- Data types for energy consumption modelling in the buildings [24].

In another description, data-driven models are divided into two categories: black-box models and gray-box ones. Black-box models usually use condition data as inputs and building energy consumption or any other observable parameter of the operation as the output. Therefore, black box models are not usually informed about characteristics of the building like the envelope or the mechanical system. On the other hand, gray-box models may use some of the building characteristics along with the condition data that would be used as inputs [24], [25]. White box models, also known as numerical or physics-based models, aim to consider all effective parameters on the intended output. These models can be more complicated, time-consuming, and expensive compared to other modelling approaches [26]. Pros and cons for different modelling methods are presented in Table 2-2.

In recent years, the widespread deployment of building management systems has produced large volumes of high-resolution operational data. At the same time, advances in data mining and machine learning have improved the ability to extract patterns from such datasets. Together, these developments have accelerated the adoption of ML-based models for predicting building energy performance [27].

Table 2-2-Advantage and disadvantages of the building modelling approaches [24], [26].

Approaches	Advantages	Disadvantages
Black box	✓ Very short development time. ✓ When the training data exhibits sufficient quality, it can serve as a precise indicator of building performance.	✗ Acquiring appropriate training data can incur significant costs. ✗ The absence of a connection between building design parameters and output data. ✗ Any alteration in building characteristics necessitates the acquisition of new input data, leading to the need for re-training.
Gray box	✓ Reduced development time when compared to numerical simulations. ✓ Offers precise building performance predictions when high-quality input data is available. ✓ Establishes a connection with certain physical building parameters.	✗ High level understanding of statistical models and engineering models along with good knowledge of building systems is required. ✗ We can only extend the model to incorporate the potential for combining or simplifying input parameters. ✗ If alterations occur in building materials, schedules, or operations, re-training becomes necessary.
Simulation (Detailed)	✓ Presents output for building energy performance with detailed parameters ✓ Provides a link for energy performance with physical building systems and condition data.	✗ Over parametrized and under determined energy performance prediction. ✗ High order required time ✗ Expensive and needs high level expertise

2.2.1 Clustering and Regression Approach

Data-driven approaches for building energy prediction range from simple regression models to complex deep neural networks and reinforcement learning controllers. While advanced models may achieve high predictive accuracy, they often operate as black boxes, limiting interpretability and reducing transparency in engineering decision-making. In building performance analysis, understanding the relationship between input variables and predicted outputs is often as important as predictive accuracy itself [28], [29], [30]. Furthermore, highly flexible models may be prone to overfitting, where strong training performance does not generalize to unseen operational conditions.

Clustering methods have been widely applied in building energy studies to identify recurring consumption patterns and segment buildings or operational days into groups with similar characteristics [31]. Although various clustering techniques exist, including hierarchical clustering [32], Gaussian mixture models [33], density-based methods such as DBSCAN [34], and self-organizing maps [35], k-means clustering remains one of the most commonly adopted methods in energy applications due to its simplicity, computational efficiency, and scalability [36], [37], [38].

In the context of building energy analysis, k-means clustering partitions data into k groups by minimizing within-cluster variance, enabling identification of representative daily load patterns or operational regimes. Its direct assignment of observations to clusters supports interpretability and facilitates subsequent regression modelling [31], [39]. For time-resolved energy datasets, this characteristic is particularly useful for distinguishing seasonal and operational behaviors.

The clustering–regression framework leverages these properties by developing regression models within each identified cluster. By modelling relatively homogeneous subsets of data, prediction errors can be reduced compared to a single global model, while maintaining clarity in how energy behavior differs across operational regimes [40], [41]. This balance between simplicity, interpretability, and predictive adequacy makes the clustering–regression approach suitable for control-oriented and diagnostic applications in high-performance buildings.

2.3 Building Energy Simulation as the Basis for Surrogate Modelling

Building Performance Simulation (BPS) models are physics-based tools used to predict thermal and energy behavior of buildings under varying design and operational conditions. Unlike purely statistical approaches, BPS models rely on fundamental heat-transfer and thermodynamic principles to compute zone-level and whole-building energy performance [42], [43]. These models allow designers and researchers to evaluate energy-saving strategies, assess system performance, and explore alternative design configurations prior to construction or retrofit [24], [44].

Whole-building simulation tools perform time-step heat-balance calculations that account for building geometry, envelope properties, internal loads, HVAC systems, occupancy, and weather conditions [24]. Among available BPS engines, widely used platforms include DOE-2 [45], TRNSYS [46], and ESP-r [47]. However, EnergyPlus has emerged as one of the most commonly adopted simulation engines due to its detailed HVAC modelling capabilities, flexible system representation, and support for sub-hourly time steps [8], [48].

EnergyPlus is particularly well-suited to this purpose, given its ability to represent multi-zone thermal interactions, detailed system controls, and dynamic external inputs makes it suitable for producing high-resolution hourly energy profiles consistent with the time-resolved objectives of this thesis.

While physics-based simulation models provide detailed and physically interpretable results, they are computationally intensive, particularly when applied to large parametric studies or optimization workflows [24]. This computational burden limits their direct applicability in rapid scenario exploration or real-time decision-support contexts.

2.4 Surrogate Modelling

2.4.1 Introduction to Surrogate Modelling

The computational cost of repeated physics-based simulation creates a practical barrier to large-scale parametric exploration and design optimization. Surrogate models address this directly by replacing high-fidelity simulation with a fast, trained approximation, enabling the kind of systematic design-space coverage that would otherwise be computationally prohibitive [49].

Surrogate models, also known as meta-models, are computationally efficient approximations of high-fidelity physics-based simulations. Rather than running full building energy simulations for each design scenario, a surrogate is trained on a structured dataset of simulation inputs and outputs, and subsequently used to predict performance across the broader design space at a fraction of the computational cost [50], [51], [52].

2.4.2 Model Development Process

The process for development of a surrogate model initiates with defining the problem and parameters, then, the designer creates an initial model and selects samples for simulation. Simulation results are stored in a database. Subsequently, a surrogate model is developed from the input-output data, and its accuracy is validated by comparing surrogate predictions to simulation outcomes for the same inputs. Surrogate derivation commonly follows a sequential approach, where sample locations are first generated using a Design of Experiments (DoE) strategy, and then the surrogate model is fitted. This static sampling

approach entails defining samples before simulation, without adjustments based on model outcomes. In contrast, the iterative approach integrates sample definition and surrogate model fitting. Samples are added iteratively to the database based on surrogate predictions and simulation results. This approach evaluates surrogate accuracy, design space complexity, or an optimization criterion to identify optimal choices for additional samples [53], [54], [55], [56]. A schematic chart for illustrating surrogate modelling development is presented in the Figure 2-2.

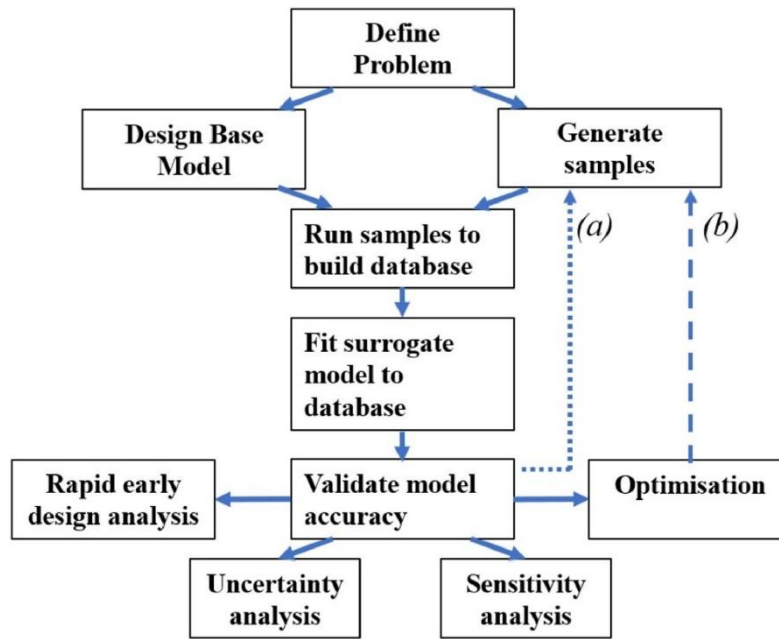


Figure 2-2- Summary of the stages in generating a surrogate model and its applications. Adopted from [56]

2.4.3 Sampling Strategies and Dataset Size

The development of surrogate models for building energy analysis relies fundamentally on two interrelated considerations: the strategy used to generate training samples and the size of the dataset required to ensure predictive reliability. Sampling approaches designed to explore multidimensional design spaces can generally be classified as static or adaptive. Static methods include pseudo-random techniques such as Monte Carlo sampling, quasi-random sequences such as Sobol methods, and stratified approaches such as Latin Hypercube Sampling (LHS), the latter being the most widely adopted in building energy applications due to its ability to provide uniform coverage of the parameter space with relatively high efficiency. In contrast, adaptive sampling strategies such as Expected Improvement criteria within Efficient Global Optimization frameworks, variance-based (uncertainty) sampling, and sequential Kriging refinement methods, iteratively refine the selection of sample points based on model performance, aiming to improve accuracy by balancing exploration and exploitation of the design domain [51], [57], [58], [59]

Closely linked to the choice of sampling method is the question of dataset size, which remains a persistent challenge in surrogate modelling for building energy consumption. While surrogate models offer a computationally efficient alternative to physics-based simulations, their predictive capability depends strongly on the number and diversity of training samples. Although formal, standardized guidelines for determining sample size remain limited, several studies have addressed this challenge through systematic empirical evaluation by testing multiple sample sizes and comparing prediction errors. This approach provides researchers with a defensible basis for sample size selection based on observed model convergence behavior rather than arbitrary judgment [18], [60]

Birdsell and Evins [61] developed surrogate models based on different machine learning models (Gaussian Process, Support Vector Machine, Gradient Boosted Trees, and Multi-Layer Perceptron) for building design problems using 1,040 training samples. Beccali et al. [62] successfully trained artificial neural networks on approximately 600 datapoints for non-residential building energy performance assessment in Southern Italy. These smaller-scale studies demonstrate that successful surrogate modelling can be achieved with relatively limited datasets when the problem complexity is manageable, the input parameter space is not excessively high-dimensional, and the target outputs are well-defined continuous variables such as energy consumption or heating/cooling loads. The relatively simpler model structures in these cases, combined with focused prediction targets, allowed the algorithms to learn the underlying relationships effectively even with fewer training samples compared to more complex, multi-output surrogate models. [62]

Vallianos and Kapsis [60] employed 3,000 samples for surrogate models corresponding to single HVAC systems and 5,000 samples for surrogate models with HVAC systems as categorical inputs. This difference highlights that adding HVAC system complexity increases the required sample size. Similarly, Shirzadi et al. [18] conducted comprehensive analyses examining how model accuracy changed with varying sample sizes for building energy and cost prediction, finding that prediction accuracy stabilized around 2,000 samples for most applications. Both studies reached similar conclusions: approximately 2,000 to 3,000 samples provide an optimal balance between computational efficiency and predictive accuracy for building energy surrogate modelling applications. Both sets of studies targeted low-dimensional outputs such as annual EUI and cost, and their sample size findings may not directly translate to surrogate models predicting high-dimensional temporal outputs such as hourly sequences.

Lopes and Lamberts [63] used approximately 244,000 datapoints but demonstrated that datasets as small as 36,600 datapoints provide equivalent accuracy, highlighting potential oversampling in many studies. Thrampoulidis et al. [12] utilized data from approximately 13,000 buildings for retrofit analysis across Zurich. Thrampoulidis et al. [64] extended their work to cover all of Switzerland with 300 archetype

models for 53 climate regions, resulting in 15,900 total combinations. Similarly, Zhang et al. [65] created datasets of approximately 10,000 retrofit scenarios for energy retrofit decision-making. Ferreira et al. [66] employed a large-scale simulation dataset of 120,000 EnergyPlus runs (10,000 variants for each of 12 midrise office archetypes) to develop an unsupervised learning-based surrogate modelling approach. Rather than using the dataset for direct input-output prediction, they clustered load signatures derived from the simulations and associated each cluster with a set of plausible envelope, internal gain, and HVAC operational features. Tested on five buildings unseen by the algorithm, the method was able to identify building features with reasonable accuracy across all archetypes, demonstrating how large synthetic datasets can enable pattern discovery and feature estimation even when direct inverse modelling is challenging. Finally, Ali et al. [67] employed Monte Carlo analysis with 10,000 simulations to generate training datasets for surrogate models predicting building energy consumption and peak loads in the Gulf Cooperation Council region. The study used 16 operational parameters with normal and triangular distributions to generate four energy intensity outputs. Extreme Gradient Boosting (XGBoost) achieved R^2 values as high as 0.99 in predicting these energy performance metrics. These studies collectively confirm that required sample size scales with problem complexity and output dimensionality.

2.4.4 Energy Output Resolution

The temporal resolution of energy predictions is a critical dimension in surrogate modelling for building energy analysis. Surrogate models, trained on the inputs and outputs of physics-based building performance simulation (BPS) tools like EnergyPlus or OpenStudio, can generate predictions at various temporal granularities. These range from annual and monthly to daily and hourly resolutions. Each temporal resolution serves distinct analytical purposes, from high-level policy analysis to detailed grid interaction studies, and presents unique challenges regarding model complexity, computational requirements, and the necessary sample size of unique building designs for adequate training. This literature review examines the current state of research on temporal resolution in building energy surrogate modelling, specifically focusing on studies where surrogate models are trained on data generated from physics-based simulation tools.

2.4.4.1 Annual and Monthly Resolution

Annual resolution represents the most extensively explored temporal granularity in building energy surrogate modelling. A significant portion of existing research focuses on predicting annual energy consumption or EUI. This prevalence reflects the computational efficiency of training models on annual totals and their suitability for high-level applications such as energy benchmarking, building stock modelling, and preliminary policy analysis. For instance, Thrampoulidis et al. [12], [64] developed surrogate models to approximate optimal building retrofit solutions, using annual energy savings as a key

performance metric. Similarly, Zhang et al. [65] employed ANNs to review and predict annual energy performance, while Beccali et al. [62] used similar techniques to assess the annual energy performance of a non-residential building stock. Shirzadi et al. [18] also contributed to this domain, developing and comparing surrogate models (RF, XGBoost, MLP) trained on 2000 simulations to predict both annual EUI and costs for building design, demonstrating high accuracy and significant computational speedups. Birdsell and Evins [61] evaluated the performance of various surrogate modelling techniques for building design problems, typically framed around annual objectives. Furthermore, Vallianos and Kapsis [60] developed highly accurate multi-output ANN surrogate models to predict annual EUI, TEDI, and CEDI for evaluating HVAC retrofit options in multi-unit residential buildings.

Monthly resolution has also received considerable attention, offering a balance between computational simplicity and the ability to capture seasonal variations. This granularity is particularly useful for retrofit analysis where performance across heating and cooling seasons is important. The work by Lopes and Lamberts [63] exemplifies this approach, developing a metamodel to predict monthly cooling energy consumption for HVAC systems in office buildings across different climates. The follow-up study by Thrampoulidis et al. [64] on large-scale retrofit analysis also incorporated monthly resolution to better capture the time-varying effects of energy efficiency measures.

2.4.4.2 Daily and Hourly Resolution

In contrast to annual and monthly models, higher temporal resolutions like daily and hourly remain significantly less explored. The increased data volume and model complexity required to capture diurnal variations and short-term operational dynamics have historically limited their development.

Markarian et al. [68] provide a critical investigation that directly addresses this gap, developing surrogate models for multiple resolutions including daily and hourly. Using 100 office building configurations, they demonstrated that hourly and daily models achieved high accuracy (adjusted $R^2 > 0.90$), confirming the feasibility of high-resolution surrogate modelling. Notably, they found that models trained on the same configurations to predict annual totals performed significantly worse, highlighting how high-resolution targets can provide richer learning signals.

Among the few studies targeting hourly resolution, Markarian et al. [19] further contributed to this area by developing surrogate models to predict hourly thermal comfort alongside annual electricity consumption and peak loads, achieving an adjusted R^2 of 0.99 for hourly predictions using 500 building simulations.

Govindarajan et al. [69] developed surrogate models for hourly energy performance prediction for high-density residential precincts, employing 2,000 unique designs to generate 17.5 million hourly records.

This study underscores that while hourly modelling is feasible, it often requires a commensurately large training dataset to capture urban-scale complexity.

2.4.5 Surrogate Modelling Applications

Surrogate models have been applied to various aspects of building design, with the most common being for:

- **Energy demand estimation (Conceptual design stage):** Surrogate models have been widely applied in early-stage building design to approximate the outputs of high-fidelity simulations and enable rapid comparison of design alternatives [70]. Prior studies have developed surrogates for a range of performance targets, most commonly aggregated annual metrics such as total energy use or EUI [71], as well as heating and cooling loads [72]. A smaller body of work has extended surrogate prediction to higher-resolution outputs, including daily electricity consumption [73] or other time-resolved demand measures. Collectively, this literature shows that surrogate models are effective for accelerating design-space exploration, although the choice of target output (annual indicators versus time-resolved profiles) strongly influences both model complexity and data requirements.
- **Sensitivity Analysis:** Surrogate models can be particularly useful for sensitivity analysis because they reduce the computational burden associated with repeatedly evaluating high-fidelity simulations. Sensitivity methods are commonly grouped into local and global approaches. Local methods examine the effect of small changes in input variables around a specific point, while global methods assess the impact of input variations over the entire input space [74]. Both local and global methods rely on simulation samples. Fast surrogate model evaluations can expedite sample generation, especially for variance-based methods that require numerous samples. For instance, calculating Sobol indices is sample-intensive and often restricted to a few parameters due to high computational costs. Surrogate models, being faster, allow for the analysis of more parameters, overcoming this limitation [75], [76].
- **Building design optimization and retrofits:** Surrogate models can reduce the computational burden of Building Design Optimization by replacing repeated high-fidelity simulations with fast approximations during the search process. In practice, surrogates are used in two primary ways. In direct surrogate-based optimization, a surrogate is trained on an initial set of simulation results and then used as the objective function for optimization. In iterative (adaptive) optimization, the workflow alternates between updating the surrogate and adding new simulation points in regions that are predicted to be promising, improving accuracy where it matters most for the optimum

[77], [78]. Surrogate models accelerate the assessment of retrofit parameters compared to conventional simulations, thereby enabling more efficient and effective building retrofits [79].

- **Fault Detection and Diagnostics:** Gunay et al. [80] introduced an inverse grey box model-based method for fault impact analysis in HVAC systems, using virtual meters derived from building automation system data to estimate energy savings from fault correction without requiring custom physics-based models.

2.4.6 Surrogate Model Types

Several types of surrogate models have been discussed in the literature but they can be categorized into two major groups: Parametric and non-parametric models [56]. Parametric models and non-parametric models differ in how they approach the relationship between inputs and outputs. Parametric models assume a specific form for this relationship and derive a data model based on that assumption. The parameters of this model are then adjusted using the collected data. In contrast, non-parametric models do not start with a predefined data model. Instead, they aim to identify the underlying relationship between the inputs (X) and outputs (y) directly from the data, without assuming a specific form[81].

Non-parametric models such as Gaussian Processes (GP) [82], Bayesian Deep Learning (BDL) [83], Artificial Neural Networks (ANN) [79], [84], Convolutional Neural Networks (CNN) [85], and K-Nearest Neighbors (KNN) [72] have been commonly employed for their flexibility and accuracy. However, these models can be complex and computationally intensive to train. Parametric models offer simpler alternatives. For instance, Carreras et al. [77] used cubic spline interpolation for optimizing insulation thickness in building designs, achieving significant time savings compared to simulation-based methods. Regressions are the most popular type of parametric models used for surrogate modelling [70], [71], [86].

Although each model type has its own advantages and disadvantages depending on the specific modelling requirements, many experts recommend initially using multiple models to determine which one is most suitable [70], [87].

2.4.6.1 ANN (MLP) and CNN Surrogate Models

Bhosekar and Lerapetritou [57] provide a comprehensive review of surrogate-based modelling advances, highlighting that ANNs offer superior flexibility in capturing nonlinear input-output relationships compared to polynomial-based methods, making them particularly suitable for complex building energy systems.

Multilayer Perceptrons (MLPs) have been extensively applied as building energy surrogates due to their universal approximation capabilities. Hou and Evins [84] developed a systematic protocol for developing

and evaluating neural network-based surrogate models, demonstrating that properly configured MLPs can achieve high prediction accuracy while reducing computational time by orders of magnitude compared to detailed simulation tools like EnergyPlus. The Rectified Linear Unit (ReLU) activation function has become the predominant choice in these architectures due to its computational efficiency and ability to mitigate vanishing gradient problems during training [29]. Similarly, Nutkiewicz et al. [88] employed data-driven MLPs with ReLU activations for urban building energy modelling, showing effective scalability across building stocks.

Ferreira et al. [89] demonstrated the application of ensemble ANNs as inverse surrogate models, training them on three-parameter change point models (3P CPMs) derived from EnergyPlus simulation data to predict 15 building features numerically. Their work highlights the flexibility of ANNs in capturing complex, nonlinear relationships between load signatures and building characteristics, while also revealing challenges in predicting certain envelope features simultaneously due to correlations in the input data.

Convolutional Neural Networks (CNNs) have gained traction for capturing temporal patterns in building energy data. Westermann et al. [85] demonstrated that deep temporal convolutional networks (TCNs) can effectively serve as surrogate models spanning multiple climate zones, leveraging ReLU-activated convolutional layers to process sequential data with varying temporal receptive fields. This approach addresses limitations of traditional MLPs in handling time-series building loads. Kim et al. [90] further validated CNN-based approaches with ReLU activations for predicting hourly cooling loads with improved generalization.

Furthermore, the interpretability of neural network surrogates remains an active research area. Rieger et al. [4] proposed structured neural network architectures that enhance explainability, a critical consideration for building practitioners who require physical insight into model predictions. Recent work by Westermann and Evins [56] addressed uncertainty quantification in neural network surrogates using Bayesian approaches.

Comparative studies suggest that while MLPs excel at steady-state energy predictions, CNNs and their variants (TCNs, LSTMs) demonstrate superior performance for dynamic load profiles [91]. The choice between architectures ultimately depends on the specific application requirements, including temporal resolution, training data availability, and interpretability needs.

2.4.6.2 Hybrid Surrogate Models

Beyond buildings, hybrid neural-network architectures that combine different types of layers are getting popular because they can simultaneously process diverse inputs and capture complex patterns in the

outputs [92]. This section provides a brief overview of such hybrid models, with a particular interest in architectures that integrate multiple model types, such as MLPs, CNNs, and tree-based methods. A primary strategy is to combine deep learning components (like CNNs for spatial features or LSTMs for temporal sequences) with other statistical or machine learning methods to improve robustness or interpretability. Examples of Hybrid Models Incorporating Diverse Algorithms include but not limited to:

- **Attention-based LSTM-MLP Fusion:** Fañanás-Anaya et al. [93] introduced a hybrid MA-LSTM-MLP architecture for dynamical system simulation. This model integrates Multi-Head Attention (MHA) mechanisms, Long Short-Term Memory (LSTM) networks, and MLPs to model complex nonlinear dynamics, demonstrating the fusion of sequential processing and feature learning. The model was trained using the Adam optimizer.
- **CNN-LSTM for Reduced-Order:** Nakamura et al. [94] developed a CNN-LSTM-based reduced-order model (ROM) for simulating three-dimensional complex flows. This architecture uses a CNN encoder to extract spatial features from flow field snapshots and an LSTM network to learn and predict the temporal evolution of these encoded features, showcasing a sequential hybrid approach for spatiotemporal data. This sequential hybrid approach for spatiotemporal data contrasts with the multi-input multi-output (MIMO) recurrent strategies evaluated by Fan et al. for building energy prediction [95].
- **MLP and Gradient Boosting:** Madaeni et al. [96] developed a model that uses an MLP and an Extreme Gradient Boosting (XGBoost) model in a complementary framework for river ice jam prediction, illustrating a hybrid between neural networks and tree-based methods.
- **Attention-based CNN-LSTM-XGBoost:** Shi et al. [97] proposed a hybrid model that combines an attention-based CNN-LSTM network with XGBoost for stock price prediction, showcasing a three-way blend of deep feature learning, temporal attention, and powerful gradient-boosted tree ensembling.
- **Bayesian and Neural Network Hybrids:** Li et al. [98] developed multi-output models for building energy prediction by combining a Bayesian adaptive spline surface (BASS) model with a Deep Neural Network (DNN), using the BASS model for global trend capture and the DNN for modelling complex residuals. This represents a hybrid of a statistical model and a neural network.
- Feng et al. [99] proposed a CNN-Transformer surrogate for geological carbon sequestration, combining convolutional layers with attention mechanisms for complex spatiotemporal modelling.

A common and effective pattern in hybrid deep learning is the fusion of MLPs and CNNs. This leverages the MLP's strength in learning from structured, tabular data and the CNN's proficiency in extracting local patterns from grid-like data (e.g., time series, images).

- **Parallel Fusion Models:** Several studies employ a parallel or early-fusion approach. Li et al. [92] developed a deep learning CNN-MLP hybrid fusion model for predicting the mechanical properties of carbon fibers. Their model processes visual microstructure data (as two-point statistics images) with a CNN branch and textual property data (macroscopic composite and matrix properties) with an MLP branch, fusing these multimodal features for final prediction. The model utilized ReLU activations and the Adam optimizer. Similarly, Park et al. [100] configured a hybrid MLP-CNN model for ground-sink susceptibility prediction. This model is designed to handle multisource geospatial data, where an MLP processes tabular factors (e.g., slope, geology) and a CNN concurrently processes spatial raster data (e.g., remote sensing imagery), with their learned features fused to produce a susceptibility map.
- **Sequential MLP-to-CNN Models:** Other works use a sequential structure where an MLP processes inputs before a CNN generates outputs. Zhao et al. [101] developed a model for district heating load prediction that uses dense layers (MLP) for initial feature extraction from input parameters, followed by 1D convolutional layers to capture temporal patterns in the output sequence. Their convolutional layers used a kernel size of 3 and ReLU activation, and the model was trained with the Adam optimizer.
- **Temporal Convolutional Networks:** Westermann et al. [85] demonstrated that a pure Temporal Convolutional Network (TCN), a specialized 1D CNN with dilated convolutions, can serve as an effective building energy surrogate, validating the use of convolutional architectures for temporal output generation.
- **Encoder-Decoder Frameworks:** The encoder-decoder paradigm, effective for sequence-to-sequence tasks, is widely used. Du et al. [102] and Chandra et al. [103] systematically evaluated such frameworks for time series forecasting. While not all use CNNs, this architecture is a common shell for hybrid components. Yang et al. [104] proposed a Temporal Feature Decomposition Fusion Network (TFDFNet) with an encoder-decoder structure using convolutional layers for building energy prediction. Their network used LeakyReLU activations and was optimized with Adam.

2.4.7 Activation Functions, Optimizers, and Kernel Sizes

The technical implementation choices in hybrid surrogate models show strong convergence, which informs best practices. When it comes to activation functions the Rectified Linear Unit (ReLU) and its variants (LeakyReLU, Parametric ReLU) are the predominant choices in both MLP and CNN components due to their simplicity and effectiveness in mitigating the vanishing gradient problem [4][5][92]. For instance, Yang et al. (2024) used LeakyReLU in their TFDFNet [104], while Zhao et al. (2022) and Li et al. (2023) employed ReLU [92], [101]. The use of tanh activation in LSTM cells is also common for capturing temporal dependencies [93].

Regarding optimizers, The Adam is the de facto standard for training these hybrid models [92], [93], [101], [104], [105]. Its adaptive learning rate and momentum properties make it well-suited for the complex, non-convex loss landscapes of deep hybrid networks. Variants like AdamW are sometimes used to improve generalization with weight decay.

Finally, in 1D CNNs used for temporal surrogate modelling, common kernel sizes are 3 or 5 [85], [101]. A smaller kernel (e.g., 3) captures short-term dependencies efficiently, while a slightly larger kernel (e.g., 5) captures broader temporal patterns. This choice is often validated through ablation studies [101]. For 2D CNNs processing spatial or image data like microstructures, typical kernel sizes are 3×3 or 5×5 [92].

2.4.8 Advantages of Surrogate Modelling

While the advantages of surrogate modelling are well-documented, their realization for time-resolved predictions remains particularly challenging. Surrogate models are faster and significantly more cost-effective to develop after data collection. More specifically, hybrid ML models offer a powerful and flexible framework for surrogate modelling by leveraging the complementary strengths of different algorithms. The literature shows successful integration of CNNs and MLPs in both parallel [92] and sequential [101] manners, as well as the incorporation of other methods like LSTMs [93], [94], attention mechanisms [93], [99], and Bayesian models [98]. While a standardized "MLP encoder with a hybrid MLP+CNN decoder" is not yet a widely reported archetype, the existing patterns provide a strong foundation for mapping scalar inputs to complex outputs. The consistent use of ReLU-family activations, the Adam optimizer, and small convolutional kernels forms a robust technical basis for model development across diverse engineering and scientific domains, from materials science [92] to fluid dynamics [94] and building energy [85], [101], [104].

The advantages of surrogate modelling across four key areas of the building design process are presented in Table 2-3, based on several studies that have explored the application of this method [55], [56], [106]. While using data generated from building simulations for surrogate modelling is feasible, access to such data remains a challenge for researchers.

Table 2-3- Overview of the benefits associated with using surrogate models in the building design [56].

Design stage	Advantages of surrogate model
Preliminary Design and Concepts	• Quick assessment of design concepts • Facilitates exploration of diverse design alternatives and scenarios within practical timeframes. • Quickly assess how design decisions impact the variability in design outcomes • Time saving by allowing for faster decision-making and iteration in comparison to traditional time-intensive simulations.
Variability Assessment of the Parameters	• Facilitate sensitivity analyses, allowing designers to identify key parameters influencing building performance efficiently. • Lower computational time and cost in parameter analysis.
Uncertainty Analysis	• Rapid derivation of the probability distribution for building performance.
Optimization Process	• Speeding up the optimization process. • Facilitating gradient-based optimization

A web-based tool developed by Ferreira et al. [107] illustrates several key advantages of surrogate modelling for building energy analysis. By enabling rapid portfolio-level retrofit screening from only heating and cooling load data, their ensemble ANN model demonstrates: (1) quick assessment of retrofit potential without detailed on-site audits, (2) facilitation of diverse scenario exploration across large building stocks, (3) significant time savings compared to traditional simulation-based approaches, and (4) acceleration of the retrofit prioritization process for building owners and managers, directly supporting the benefits outlined in Table 2-3. Together, these applications establish the methodological foundations upon which the surrogate modelling framework developed in this thesis builds.

2.4.9 Research Gaps

Despite the increasing use of surrogate modelling in building performance research, several important gaps remain, particularly in relation to high-performance buildings and time-resolved analysis.

Gap 1: Limited focus on high-performance and net-zero buildings

A large portion of the existing surrogate modelling literature focuses on conventional buildings and standard design cases [56], [67] High-performance and net-zero energy buildings, which involve more complex interactions between envelope, systems, and operation, remain comparatively underexplored. This limits the applicability of current surrogate models in advanced design contexts where accurate representation of these interactions is critical.

Gap 2: Predominance of low-resolution performance metrics

Most surrogate models in the building domain are developed to predict aggregated outputs such as annual or monthly energy consumption [56], [91]. While suitable for benchmarking and early-stage assessments, these metrics do not capture the temporal dynamics required for detailed performance evaluation, control strategy analysis, and grid-interactive applications. High-resolution surrogate models capable of predicting hourly performance remain limited in the literature [85], highlighting a significant gap.

Gap 3: Limited consideration of design parameters in high-resolution models

Due to the relatively small number of studies on hourly surrogate modelling, many input parameters, particularly geometric and design-related variables, are either simplified or excluded [72], [85]. In several cases, parameters are represented using discrete or highly constrained values, which restricts the ability of surrogate models to support comprehensive design exploration. This indicates a need for models that can incorporate a broader and more continuous range of building design variables.

Gap 4: Lack of structured linkage between physics-based simulation and surrogate modelling

Although surrogate models are typically trained on simulation-generated data, the process of integrating physics-based models, sampling strategies, and machine learning techniques is often not presented as a cohesive framework [70], [84]. This results in limited transparency regarding how datasets are generated and how surrogate models can be systematically applied within the building performance simulation (BPS) workflow.

These gaps collectively motivate the surrogate modelling framework developed in Chapter 3, which addresses high-resolution temporal prediction for high-performance office buildings through a structured LHS-driven simulation and deep learning pipeline.

Chapter 3

3 Methodology

3.1 Overview

This chapter presents the methodological framework developed to investigate the energy behavior of high-performance office buildings and to design a surrogate model capable of predicting time-resolved energy performance under Waterloo's climate. The framework is designed to address the gaps identified in Chapter 2, specifically the limited focus on high-performance buildings, the predominance of low-resolution surrogate outputs, the restricted treatment of design parameters in hourly models, and the absence of structured physics-to-surrogate pipelines. To address these gaps, the framework integrates measured data analysis, physics-based simulation, and machine learning into a unified workflow that supports the development of predictive models from design-stage inputs.

The research is structured into four main phases, as illustrated in Figure 3-1. These phases are sequentially connected, where the outcome of each phase directly informs the subsequent stage of the methodology. In addition to supporting the next phase, outputs from earlier stages are also reused for model validation, parameter selection, and interpretation of surrogate model results, ensuring consistency between empirical observations, simulation-based modelling, and data-driven predictions.

- **Phase 1: Case study data analysis.**

Measured data from a real-world high-performance office building are analyzed to develop an initial understanding of building energy behavior. Clustering and regression techniques are used to explore temporal patterns and relationships between key variables. The primary objective of this phase is to identify influential parameters and characterize their behavior, thereby informing the selection of input variables for the parametric modelling framework developed in Phase 2. In addition, the insights derived from this phase are used beyond parameter selection. They support the validation of the parametric numerical model developed in Phase 2 by providing a reference for expected building behavior under real operating conditions. Furthermore, the clustering-based representation of energy profiles developed in this phase is later extended as an interpretive tool for analyzing surrogate model outputs, enabling a structured understanding of predicted hourly energy behavior at the annual scale.

- **Phase 2: Development of a parametric numerical model.**

A physics-based parametric energy model is developed to represent high-performance office buildings in Waterloo, Ontario. Based on insights from Phase 1 and relevant literature, key design

parameters are selected and varied within defined ranges. This phase establishes a systematic simulation environment capable of capturing the relationship between design variables and building energy performance, forming the basis for dataset generation in Phase 3. The performance of the parametric model is evaluated against insights derived from the case study data to ensure that the simulated behavior remains consistent with observed building operation, providing confidence in its use for synthetic dataset generation.

- **Phase 3: Synthetic dataset generation and preprocessing.**

The parametric model is used to generate a structured dataset by sampling across the defined input space. This dataset captures a wide range of design scenarios and corresponding energy performance outputs. The generated data are then processed, including normalization and partitioning into training and validation sets, to support the development of surrogate models in Phase 4.

- **Phase 4: Surrogate model development and evaluation.**

Machine learning models are developed to approximate the relationship between design parameters and building energy performance. The models are designed to predict both aggregated (e.g., annual) and time-resolved (e.g., daily) energy outputs from low-dimensional inputs. Their performance is evaluated in terms of predictive accuracy and their ability to reproduce meaningful temporal patterns, enabling efficient energy estimation and parameter impact analysis. In addition to point-wise prediction accuracy, clustering-based representations are used to interpret the predicted hourly profiles, providing an additional layer of analysis that supports engineering understanding of seasonal patterns, peak demand periods, and operational behavior.

It is important to note that this research framework is predicated on the widely accepted premise that numerical building performance simulations, although subject to inherent uncertainty when applied to unbuilt designs, remain a reliable decision-support tool for comparative analysis and high-performance building design [108], [109].

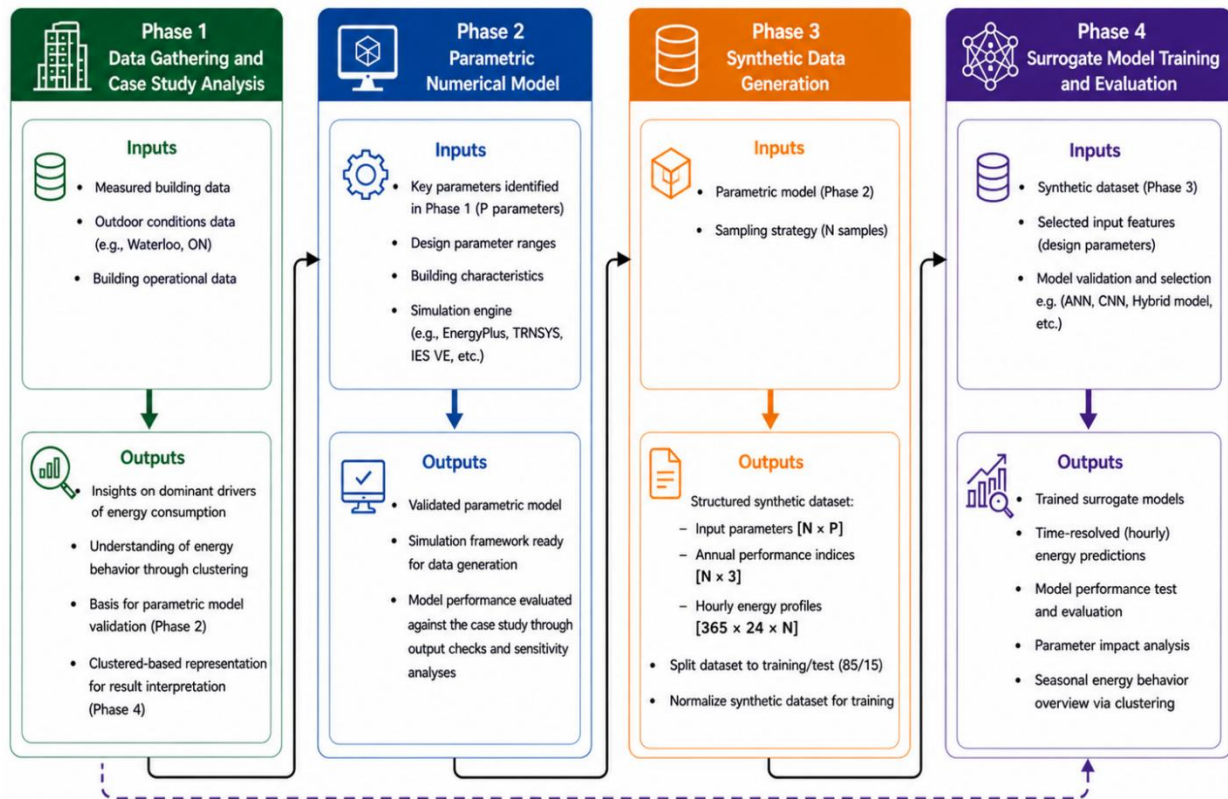


Figure 3-1- A General Four-Phase Framework for Physics-Informed Surrogate Modelling of Building Energy Performance.

3.2 Phase 1 – Data Gathering and Case Study Analysis

3.2.1 Case Study Description

The initial phase of unraveling the performance of a NZEB involves an overall examination of a real-world case study in this domain. In this research, the Office Case Study (i.e., evol1) is a net positive energy and zero carbon building that represents the cutting-edge in sustainable building design. Energy analysis of the building was conducted, encompassing critical aspects such as its building envelope, mechanical equipment and other building services. This detailed and holistic description of the Office Case Study is intended to accentuate its pivotal characteristics and guide phase 2 on the development of a parametric numerical model that can represent all NZEBs within the design input boundaries, under the Waterloo Region climate.

3.2.2 Data Extraction and Cleaning

The energy consumption data, covering January 2021 to December 2023, was collected using smart meters integrated into the BMS of the office case study. Initial data cleaning involved identifying and

addressing days with data storage errors or missing values. Each day with such errors is flagged for further investigation. Recorded days with zero energy consumption were identified and removed from the dataset. Days with zero consumption data may indicate data collection issues or non-operational periods. A total of eleven days were detected and removed from the dataset. Another step in the data cleaning would be dealing with limited missing or zero value data. For days with less than two hours of zero data (i.e., less than five days with one hour of zero energy consumption detected in considered time period), the missing hourly data was interpolated by taking the average of the energy consumption values one hour before and one hour after the missing data point. Temporal consistency of the dataset was checked by ensuring that each day contains 24 hourly energy consumption datapoints per day. It was ensured that timestamps are sequential and consistent throughout the entire dataset. The final step in data cleaning was to check the integrity of the data. Additional integrity checks were performed to identify and resolve any anomalies such as duplicates or inconsistencies in the dataset in order to achieve a clean and smooth dataset for subsequent steps [110].

3.2.3 Data Normalization

Normalization is a crucial step in preparing energy consumption data for analysis. It ensures that variables are on a consistent scale, which is essential for effective operation when applying clustering algorithms and predictive models. The Min-Max scaling method, also known as feature scaling, was used to normalize the energy consumption data. This method scales the data to fall within a specific range, typically [0, 1], preserving the data's relative relationships while transforming it to a common scale [111], [112]. The Min-Max scaling formula for a data point X is defined as:

$$X_{normalized} = (X - X_{min}) / (X_{max} - X_{min}) \quad Eq\ 3-1$$

where $X_{normalized}$ is the normalized value of the measured parameter (e.g., hourly electricity consumption); X represents the original value of the data point; X_{min} is the minimum and X_{max} is the maximum value the parameter takes within the dataset.

3.2.4 Outlier Detection and Handling

Detecting and managing outliers is crucial as they can distort analysis and modelling results. There are various methods for outlier detection which can be used. For this study, the interquartile range (IQR) method, a robust statistical approach that identifies outliers based on the spread of data, was used [113]. A total of eleven days were identified as outliers (i.e., daily energy profiles with more than five hours being outliers) and removed from the dataset. For other days that have five hours or less identified as outliers, outliers were replaced by the average energy consumption of the hour before and the hour after

the corresponding timestep. Using the mean value to handle “mild outliers” ensures data integrity while reducing the influence of extreme values [110], [114]. A total of 41 hourly datapoints were replaced. The final dataset consists of 1095 days of daily energy profiles (or 26280 hourly datapoints). Figure 3-2 presents the daily energy consumption profiles for all the datasets before and after outlier detection.

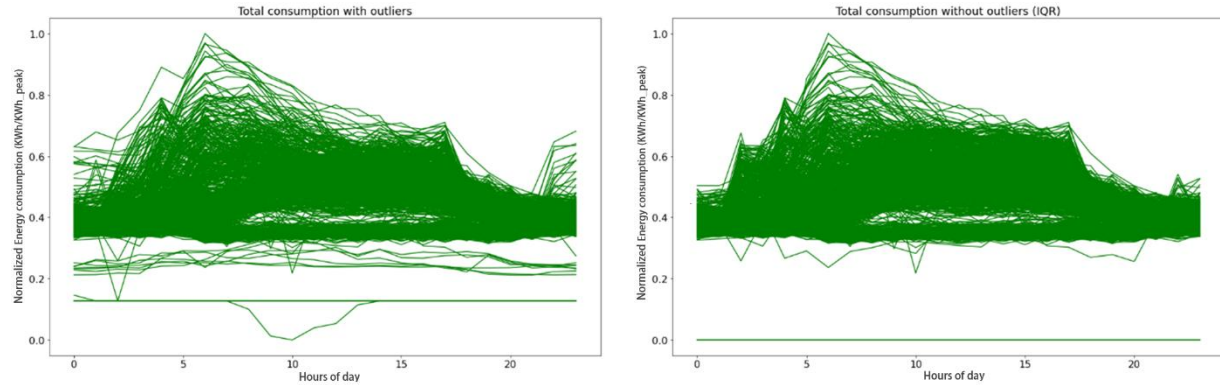


Figure 3-2- Total energy consumption before (left) and after (right) outlier detection and handling.

3.2.5 Clustering Analysis of Daily Energy Profiles

After cleaning the dataset and handling the outliers, the next step was splitting the data into suitable groups which can represent similar daily energy consumption profiles of the building. For this, the k-means clustering method was used as it has been proven to be effective and easy to use for building energy consumption applications [73], [115]. The k-means method requires two inputs: the dataset and the number of clusters. To determine the appropriate number of clusters, the Elbow method and Silhouette index can be used.

The Elbow method is a technique rooted in k-means clustering, to identify the optimal number of clusters for the dataset. By assessing the Within-Cluster Sum of Squares (WCSS) for different cluster numbers, the method determines the point at which the reduction in WCSS diminishes, forming an "Elbow" shape, which indicates the ideal cluster count. This method assists in striking a balance between cluster model complexity and separation [116]. On the other hand, the Silhouette index is a metric used to measure the quality of clustering. The index quantifies how similar an object is to its cluster compared to other clusters. The number of clusters that maximizes the Silhouette index would be the preferred number of clusters as it represents the best trade-off between cluster quality and quantity [117]. After acquiring the number of the clusters using Elbow method and Silhouette index, the k-means clustering algorithm was applied. The results of the k-means clustering were further analyzed, revealing the key characteristics and underlying patterns within each cluster as identified through unsupervised machine learning.

3.2.6 Model Prediction for Each Cluster

At this step, each cluster dataset was used to train machine learning models, aiming to predict daily energy consumption profiles of the building. While literature has indicated that advanced ML models such as ANN and Gaussian Process (GP) models can yield better predictions [56], they tend to offer reduced interpretability and user-readability. As this study aims to develop cluster models that aim to offer both interpretable and easy to understand, linear and second-order polynomial regression models were also employed. However, to gain deeper insights into feature importance, particularly regarding environmental inputs, MLP was also used. For each cluster, the dataset was split into training (80%) and testing (20%) subsets to evaluate the performance of each model. Each model was trained on the training subset and validated on the testing subset using statistical metrics such as Mean Absolute Percentage Error (MAPE), Relative Mean Squared Error (RMSE), and R-squared (R^2).

3.2.7 Clustering of Building Load Profiles

The clustering methodology described earlier was applied to individual building load components to gain a detailed understanding of the building's energy performance. By analyzing specific end uses such as HVAC, plug loads, lighting, electric vehicles (EVs) charging, and other loads, distinct operational patterns can be identified. Data insights derived from this analysis, such as occupancy schedules inferred from plug load and lighting consumption, can also contribute to the characterization and modelling of the overall building energy use. A schematic overview for the methodology is presented in Figure 3-3.

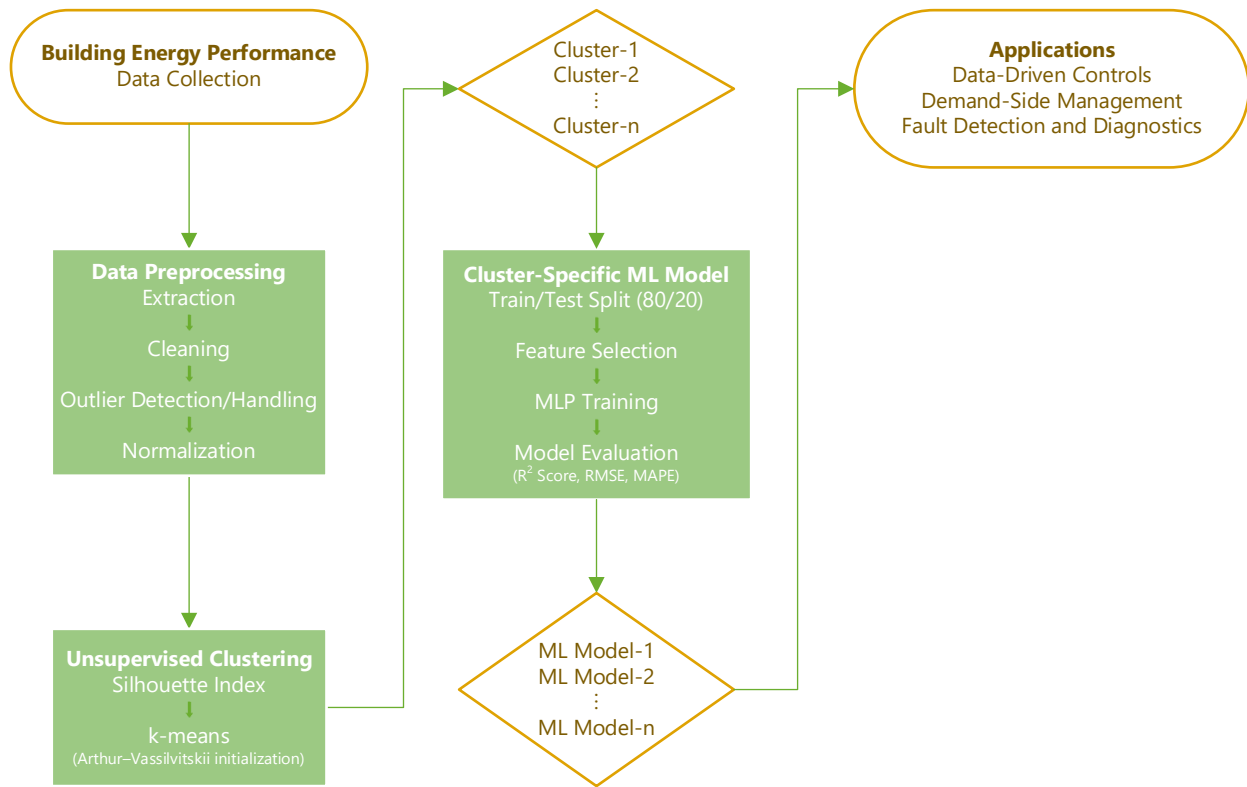


Figure 3-3- Schematic flowchart of the case study investigation using clustering regression, adapted from [118].

3.3 Phase 2 – Development of the Numerical Parametric Model

3.3.1 Simulation Tool Selection

The parametric model was developed as python script which uses EnergyPlus as the building energy simulation engine. EnergyPlus was selected due to its proven reliability, open-source flexibility, and ability to model HVAC, envelope and controls. The tool supports Python-based automation for large-scale parametric simulations, making it ideal for synthetic data generation. [119], [120], [121], [122], [123].

3.3.2 Parametric EnergyPlus Model

A physics-based building energy model was developed in EnergyPlus (v25.1) to generate the synthetic dataset used for training and evaluating the surrogate model. The model represents a generic commercial office building located in Waterloo, Ontario, Canada (43.47°N, 80.38°W) and was designed to capture the main physical processes influencing building energy consumption, including envelope heat transfer, internal gains, HVAC operation, and weather effects.

The parametric model serves as the foundation of the simulation framework. Multiple simulations are generated by systematically varying selected building design and operational parameters within predefined ranges. These simulations produce the synthetic dataset used for surrogate model development. A detailed description of the parameter selection and sensitivity considerations is provided in the Model Description.

3.3.2.1 Building Geometry and Zoning

The building geometry follows a simplified rectangular configuration consistent with typical office buildings. Each floor is divided into one core zone and four perimeter zones, with perimeter zones extending 5 m from the façade in accordance with ASHRAE zoning recommendations. This zoning approach allows the model to capture differences in solar exposure, internal gains, and HVAC responses between façade orientations while maintaining a consistent structure for parametric modifications.

3.3.2.2 Envelope and Internal Loads

The envelope constructions and internal load assumptions are based on commonly used standards for office buildings, including ASHRAE 90.1[9] and the National Energy Code of Canada for Buildings (NECB 2020) [10]. Standard assemblies were used for walls, roof, floor, and glazing systems to represent typical commercial construction.

Internal gains from occupants, lighting, and equipment follow standard office schedules and densities derived from NECB guidelines. While the schedules remain fixed, selected parameters such as lighting and equipment power densities are treated as variable inputs in the parametric simulations.

3.3.2.3 HVAC System Representation

Heating and cooling are modeled using a Variable Refrigerant Flow (VRF) system serving individual thermal zones. Ventilation air is provided by a dedicated outdoor air system equipped with an energy recovery ventilator (ERV) to recover sensible and latent heat from exhaust air. Supplementary heating and cooling coils are included to meet any remaining conditioning loads.

The HVAC system operates using dual temperature setpoint schedules reflecting typical office operation. These schedules were derived from the operational characteristics of the reference case study building and are applied consistently across all simulations.

3.3.2.4 Simulation Setup and Outputs

Weather conditions are represented using the TMY weather file for Waterloo Region, Ontario (ASHRAE climate zone 6A). Simulations are performed using 10-minute timesteps to capture building and system dynamics, with outputs aggregated to hourly resolution for surrogate model development.

The primary simulation outputs include hourly total electricity consumption and annual performance indicators such as EUI, TEDI, and CEDI. Only simulations that completed successfully and passed automated quality checks were retained in the dataset.

All simulations were executed through a Python-based automation framework using the EnergyPlus API, enabling systematic modification of model inputs, batch execution of simulations, and structured extraction of output data.

3.3.2.5 Parametric Inputs

A subset of twelve building design and operational parameters was selected as variable inputs in the parametric model. These parameters represent key factors influencing building energy performance, including building geometry, envelope thermal properties, glazing characteristics, infiltration, internal loads, and building orientation.

The selection of these parameters, along with their corresponding ranges and sensitivity considerations, is discussed in detail in the Model Description section.

3.4 Phase 3 – Synthetic Data Generation

Using the parametric building energy model, a synthetic dataset was generated through systematic variation of design inputs. The objective of this phase was to construct a diverse yet physically realistic dataset that captures the influence of variations in envelope properties, geometry and internal loads, on energy consumption. This dataset forms the basis for training and evaluating the surrogate machine learning model presented in subsequent sections.

3.4.1 Sampling Method and Size

Consistent with the sample size literature reviewed in Section 2.4.3, which identified 2,000–3,000 samples as an effective range for building energy surrogate modelling, $N = 3,000$ was selected to balance computational feasibility with sufficient coverage of the 12-dimensional parameter space for high-resolution hourly outputs. All parameters were sampled in a normalized $[0,1]$ space, after which values were mapped back to physical units and applied to the parametric energy model via the EnergyPlus Python API. A fixed random seed was used to ensure reproducibility.

Input parameters were sampled independently, without imposing predefined correlations, in order to maximize diversity in the training dataset and allow the surrogate model to learn parameter interactions implicitly. Although certain real-world correlations exist (e.g., between glazing thermal properties or geometric parameters), these relationships were intentionally not enforced at this stage to avoid restricting the exploration of the input space.

To ensure physical and geometric feasibility, parameter bounds and minimum values were defined for inputs with inherent dependencies, such as building floor area and aspect ratio, as described in the parameter variation section. During model generation, parameter combinations that resulted in invalid configurations or unsuccessful EnergyPlus simulations were automatically filtered out. Only simulations that completed successfully and satisfied all feasibility constraints were retained for dataset construction.

Sampling uniformity was verified through inspection of parameter distributions and pairwise scatter plots, confirming even coverage of the defined input ranges.

3.4.2 Data Collection and Storage

To ensure interoperability and efficient access, the simulation output data was organized into two different types:

- Annual energy metrics such as EUI, TEDI, and CEDI produce one output value per simulation. These output data are stored as a 2D matrix and exported as a CSV file.
- Hourly energy consumption for each simulation was initially stored as a $[8760 \times 1]$ vector, representing all hours in a year. For the full dataset, each vector is reshaped into a $[365 \times 24]$ matrix to capture daily profiles across the year. The results from all simulations are then combined into a 3D matrix with shape $[365 \times 24 \times \text{number of simulations}]$. This 3D matrix is saved as an .h5 file, which is easy to read and efficient for use with Python ML libraries.

3.5 Phase 4 – Surrogate Model Development and Evaluation

Following the generation of the synthetic energy consumption dataset using the parametric numerical model, the next phase of the methodology addresses surrogate model training and evaluation. This section outlines the approach used to develop, train, and evaluate surrogate models based on the synthetic dataset.

3.5.1 Data Splitting

The synthetic dataset is randomly divided into a base training set (85%) and a final test set (15%). This split is used to separate the data between initial model development and final evaluation.

In addition, a 5-fold cross-validation method is applied to the training set. This involves dividing the training data into five random subsets, where each subset is used once as a validation set while the remaining four are used for training.

After training and validating the models using 5-fold cross-validation, the best-performing model is selected. This model is then evaluated on the final test set, which was not seen during any part of the training or validation process, ensuring an unbiased assessment of its performance.

3.5.2 Model Selection

Fully connected artificial neural networks (FC-ANNs) and convolutional neural networks (CNNs) were adopted due to their suitability for different output types in this study. As literature indicates, FC-ANNs have been proven effective for predicting scalar outputs and are therefore appropriate for annual energy indicators such as EUI, TEDI, and CEDI. CNNs, by contrast, are well suited to outputs with two-dimensional structure and can capture local patterns in hourly energy data when represented as 24-hour daily profiles over a full year[29], [56], [84]. In this study, each sample is treated as a two-dimensional matrix that preserves temporal structure across hours and days.

To exploit the advantages of both architectures, a hybrid model consisting of an ANN with a convolutional output head was also developed and evaluated. This hybrid approach was intended to provide strong predictive performance for both annual energy indices and hourly energy consumption profiles.

3.5.3 Hyperparameter Tuning

To systematically optimize model performance, automated hyperparameter tuning was carried out using Optuna, a Python-based optimization framework. Table 3-1- to

Table 3-3 present the hyperparameters and their range for all three ML models. For each model, the final hyperparameter values were selected based on the best-performing configuration identified by Optuna that minimizes validation loss.

Table 3-1- Hyperparameters and their range for FC-ANN

Hyperparameter	Description	Range / Options
latent_dim	Number of neurons in each hidden layer, defining the model's layer width	64 – 512
hidden_dim	Size of the internal feature representation learned before output prediction	128 - 1024
lr	Learning rate	1×10^{-4} – 1×10^{-2}
wd	Weight decay	1×10^{-6} – 1×10^{-4}
w_hour (α)	Hourly-to-annual loss weight	0.1 – 2

Table 3-2- Hyperparameters and their range for CNN

Hyperparameter	Description	Range / Options
latent_dim	Latent dimension	128–512
conv_channels	Convolutional channels size	8-64
cnn_channels	CNN channel multiplier	32–256
lr	Learning rate	$1 \times 10^{-4} - 1 \times 10^{-2}$
wd	Weight decay	$1 \times 10^{-6} - 1 \times 10^{-4}$
w_hour (α)	Hourly-to-annual loss weight	0.1 – 2
use_2d_cnn	1D vs 2D CNN switch: Boolean (Use of 2D CNN)	True/Fulse

Table 3-3- Hyperparameters and their range for Hybrid ANN-CNN model

Hyperparameter	Description	Range / Options
latent_dim	Size of latent feature vector (model capacity)	64 – 256
conv_channels	Number of channels in convolutional head	8 – 32
lr	Learning rate for Adam optimizer	1e-4 – 1e-2
wd	Weight decay regularization	1e-6 – 1e-4
w_hour (α)	Weighting factor for hourly RMSE in total loss	0.1 – 2
Optimizer	Adaptive learning optimizer	Adam
Batch size	Fixed for all trials	64
Epochs	Training limit with patience monitoring	10 (early stopping)

3.5.4 Model Training

The model training process for each selected algorithm is applied to the dataset to determine the optimal weights. All other hyperparameters remain fixed, based on the values selected using Optuna. In the surrogate modelling approach, model weights are adjusted using the k-fold cross-validation method. For each fold, a separate model is trained, and ultimately, an ensemble model is created by combining the weights from all trained models. Using k-fold validation helps identify the best-performing model while reducing the risk of overfitting.

3.5.5 Model Evaluation

Model performance is assessed on the **unseen test set** using the following metrics:

$$\text{MAE} = \frac{1}{n} \sum |y_i - \hat{y}_i| \quad \text{Eq 3-2}$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum (y_i - \hat{y}_i)^2} \quad \text{Eq 3-3}$$

$$R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2} \quad \text{Eq 3-4}$$

Models with lower MAE/RMSE and higher R^2 are selected as final surrogates.

3.5.6 Model Selection and Comparison:

The performance of different models (ANN, CNN, and Hybrid model) is compared based on validation losses and evaluation metrics, with the model offering the best balance between predictive accuracy and computational efficiency being selected. As noted in previous sections, cross-validation techniques, such as k-fold cross-validation, are employed to ensure robustness, providing insights into model stability and performance across different subsets of the data.

3.5.7 Parameter Sensitivity and Impact Analysis

Once the surrogate model is validated, it is used to perform sensitivity analysis through parameter perturbation. Each input parameter is varied within its range while others are fixed at nominal values. The resulting changes in daily energy consumption quantify parameter influence, revealing which design variables most strongly affect building performance.

This analysis directly answers the first research question by identifying key design and operation parameters influencing high-performance office energy use.

Four complementary statistical and model-based sensitivity analysis methods were deployed. Each method was applied to different temporal resolutions of the model outputs (hourly, daily, or annual) to ensure comprehensive interpretation of parameter effects.

1. **Correlation Analysis:** A baseline linear sensitivity screening that quantifies direct relationships between each input parameter and the total daily energy consumption (calculated as the sum of 24-hour predictions). This method is applied on the annual metrics as well to provide a comparison base as well on the effect of input parameters on different output types[124], [125].

2. **One-at-a-Time (OAT) Analysis:** A local sensitivity method assessing how incremental variations in a single parameter affect the hourly energy profiles and their corresponding daily totals, while keeping other inputs fixed. The effect of each input on the annual metrics is assessed using this method as well[126], [127].
3. **Morris Elementary Effects Method:** A global sensitivity approach that captures both nonlinear behaviors and parameter interactions, applied to the daily aggregated energy outcomes to evaluate the combined effects across the input space[128].
4. **Permutation Importance:** A model-based sensitivity metric that evaluates the impact of each input feature on hourly prediction accuracy by measuring changes in model error when individual features are randomly permuted[129].

Together, these methods provide a robust framework for identifying the most influential design and operational parameters governing the model’s predictive behavior across multiple temporal scales. Table 3-4 summarizes the four sensitivity and interpretability methods applied in this study and the rationale for their inclusion.

Table 3-4- Overview of Sensitivity and Importance Analysis Methods

Method	Analytical contribution	Rationale
Correlation	Linear association	Baseline intuition, sanity check
OAT	Directional local sensitivity	Sign and magnitude of effects
Morris EE	Global + interaction effects	Nonlinearity and coupling
Permutation importance	Model-dependent accuracy impact	What the model relies on

3.6 Summary

This chapter presented the four-phase methodological framework for developing a surrogate model capable of predicting time-resolved energy profiles of high-performance office buildings. The framework integrates measured data analysis, physics-based parametric simulation, synthetic dataset generation, and machine learning into a unified workflow grounded in the gaps identified in Chapter 2. The outputs of each phase feed directly into subsequent chapters: Phase 1 clustering and empirical findings are presented in Chapter 4, Phase 2 and 3 parametric model behavior and sensitivity results are reported in Chapter 5, and Phase 4 surrogate model development, evaluation, and parameter impact analysis are presented in Chapter 6.

Chapter 4

4 Investigation of the Energy Behavior of an Office Case Study

This chapter presents the results of Phase 1 of the methodological framework, in which measured energy data from the evolvl building are analyzed to characterize temporal energy behavior and identify the input parameters that inform Phase 2.

4.1 Carbon-Neutral Office Case Study

This Chapter focuses on evolvl, a three-story office building in Southwestern Ontario, Canada, designed to be carbon-neutral and net-energy positive. The office building (Figure 4-1) has multiple tenancies, and a conglomeration of diverse businesses spanning across multiple industries.

4.1.1 Building Description

The total net floor area of the building is 9,406 m². The building is fully electric with no on-site combustion systems and located in the city of Waterloo, Ontario, Canada which has cold and humid weather classified under ASHRAE 6A climate zone (or Dfb under the Köppen climate classification). The building opened in 2019. Following some commissioning activities during that year, it has achieved and maintained operational carbon neutrality and net-positive energy performance since 2020, through the integration of energy conservation and efficiency measures as well as the incorporation of renewable energy generation [130].

4.1.1.1 Envelope Design

An essential criterion for achieving high-performance in a building is the presence of high-performance envelope that results in significant energy conservation (i.e., heating, cooling and electric lighting) and enhanced occupant comfort. The rated thermal resistance of the wall assembly is approximately $RSI_{wall} = 5.3 \text{ m}^2\text{K/W}$, while the thermal resistance of the roof assembly is approximately $RSI_{roof} = 7 \text{ m}^2\text{K/W}$. Note that these values represent the rated performance of the insulation materials only. In practice, thermal bridging, particularly from metal framing, reduces the effective thermal resistance of the full assemblies.

Windows are one of the important building envelope systems as they can exert a direct influence on space conditioning, the electric lighting energy and occupant comfort. In this case, the building incorporates triple-glazed low-emissivity windows, with a thermal transmittance (center-of-glass), solar heat gain coefficient, and visible light transmittance $U_{win} = 0.87 \text{ W/m}^2\text{K}$, $SHGC = 0.32$, and $TVIS = 0.53$, respectively. Moreover, the overall window-to-wall ratio is equal to $WWR = 0.37$. Since all the windows

of the building are non-operable (fixed), infiltration has been reduced. On the other hand, this design choice eliminates any opportunities for natural ventilation.

4.1.1.2 HVAC Design

Beyond energy conservation measures, the Office Case Study incorporates an advanced mechanical system, featuring a ground-source heat pump with an open water-loop configuration that enhances building energy efficiency. Water is drawn from an extraction well at a depth of 160 meters at approximately constant temperature of 10°C throughout the year and flows through a water-to-water heat exchanger connected to an internal water-loop. After extracting energy from the internal water loop, the water is injected into a separate well at a different temperature. Note that the internal water loop can either reject heat to, or extract heat from the geothermal loop, which will alter the temperature of the water injected into the well.



Figure 4-1- Carbon neutral office case study, located in southwest Ontario, Canada. [131]

The building employs two distinct heat pump systems: a three-way heat pump situated in the mechanical room and a Variable Refrigerant Flow (VRF) system distributed throughout the building to regulate the temperature of the various thermal zones. The three-way heat pump, which is able to generate both heating and chilled water, serves a dual purpose. Firstly, it comes into play if the heat recovery system of the air handling unit alone proves not to be sufficient. Secondly, it provides hot water for unit heaters placed at building entrances during the heating season. The VRF system, on the other hand, is responsible for heating and cooling the thermal zones. The VRF system has the ability to simultaneously cool one

zone (e.g., the server room) while transferring the extracted heat to another area of the building (e.g., perimeter zone) that requires heating. Additionally, the VRF system operates at variable capacities based on the zones' demand, allowing for energy-efficient performance. Both the three-way heat pump and the VRF system function as water-to-water heat pumps, as they interface with the internal water loop of the ground source system, facilitating the transfer of heat between the internal water loop and the fresh air supplied by the air handling unit. An electric boiler serves as a backup for the internal water loop during extremely cold days in winter, when water from the extraction loop may not provide sufficient heat. The schematic of the building mechanical system and ground-source system is shown in the Figure 4-2. The design of this HVAC system specifically the ground-source heat pump and VRF configuration, produces energy behavior that is less weather-dependent than conventional buildings, a characteristic that has direct implications for the empirical analysis presented in Section 4.2.

Figure 4-2- Simplified depiction of the building HVAC piping system (courtesy of D. Mather)

facilitated by coils, connected to the three-way heat pump. Another noteworthy feature of the building is the implementation of a transpired solar air thermal collector (TSC), supplying the AHU with pre-heated fresh air during heating season. The TSC is effective on sunny, cold winter days. Acting as a double-skin façade, it uses a fan to draw outdoor air through perforations in the cladding, pre-heating the fresh air through convection before it enters the AHU. The ventilation system is shown in the Figure 4-3.

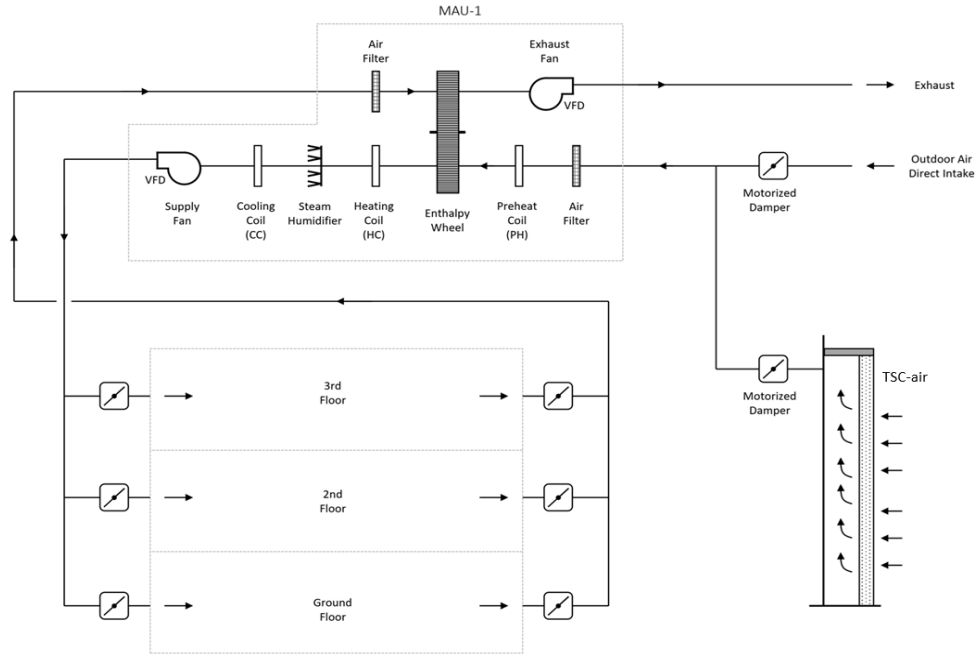


Figure 4-3-Simplified schematic of the air handling unit (AHU) system for the case study building (courtesy of D. Mather).

4.1.2 PV System Design

To achieve operational carbon neutrality, an on-site photovoltaic system is essential. The building site employs two distinct PV systems: a rooftop and a parking canopy. The rooftop PV system has a nominal generation capacity of 264 kWp, while the PV parking canopy has a nominal capacity of 504 kWp, resulting in a net-positive energy performance. As a result, 28 electric vehicle (EV) chargers have been installed in the parking lot to supply EVs with electricity (Figure 4-1). A utility net meter provides credits for excess energy exported to the grid, which are then applied when energy consumption exceeds generation (e.g., during winter).

The process of data extraction yielded a comprehensive dataset of hourly energy consumption for three years, from January 1st, 2021 to Dec 31th, 2023. This dataset, collected through 43 energy meters integrated into the Building Management System (BMS), forms the foundation of this energy analysis outlined below.

Table 4-1- Summary of site and building characteristics

Site	
Site Location	Waterloo, Southwestern Ontario, Canada
Latitude / Longitude	43.4811° N / 80.5467° W
ASHRAE Climate Zone	6A
Heating Degree Days (HDD _{18°C})	3794 (2021), 4173(2022), 3686(2023) ¹
Cooling Degree Days (CDD _{18°C})	262 (2021), 207(2022)), 150(2023) ²
Building	
Typology	Office
First Year of Operation	2019
Net Floor Area	9,406 m ²
Number of Stories	3 + mechanical penthouse
Energy Use Intensity (measured)	68.23 kWh/m ² /year (2021) 72.90 kWh/m ² /year (2022) 76.42 kWh/m ² /year (2023)
Envelope Systems	
Wall Assembly Thermal Resistance	5.3 m ² K/W
Roof Assembly Thermal Resistance	7 m ² K/W
Window-to-Wall Ratio (WWR)	0.37
Window type	triple-glazed low-emissivity
Window Thermal Conductance	0.87 W/m ² K
Solar Heat Gain Coefficient (SHGC)	0.32
Visible Light Transmittance	0.53
Solar Shading	horizontal overhangs (south) & vertical louvers (east/west)
Mechanical Systems	
Space Heating/Cooling	Open loop ground source heat pump (GSHP) with 160 m depth of extraction and injection wells Variable Refrigerant Flow (VRF) system for heating and cooling Three-way heat pump that provides additional heating or cooling Backup electric boiler
Ventilation System	Air handling unit that provides 100% fresh air and utilizes enthalpy wheel for energy recovery, with 81% efficiency.
Hot Water System	Electric boiler
Electrical Systems	
Lighting	LED, controlled by occupancy sensors; Lighting power density (LPD) of 4.75 W/m ²
Photovoltaic	264 kWp rooftop 504 kWp parking canopy
Number of Electric Vehicle (EV) Chargers	28 (AC level 2)

¹<https://kitchenerwaterloo.weatherstats.ca/charts/hdd-yearly.html>

²<https://kitchenerwaterloo.weatherstats.ca/charts/cdd-yearly.html>

4.2 Energy Behavior Analysis and Results

This section aims to analyze the energy consumption of a high-performance building and identify key parameters that: i) are essential for achieving net-zero performance targets and ii) can serve as critical inputs in developing a regression model for daily energy prediction.

The available dataset consists of high-resolution hourly energy measurements collected over multiple years, resulting in a large and complex time series that is not readily interpretable using direct analysis. To address this challenge, clustering techniques are employed to identify recurring patterns in daily energy profiles. This approach enables the reduction of data complexity while preserving key temporal characteristics, such as daily operational cycles and seasonal variations.

By grouping similar days into clusters, the building's energy behavior can be represented through a set of distinct operational regimes. This facilitates a more structured interpretation of the data and allows for the identification of patterns that may not be evident in the raw time series.

Regression models are subsequently developed within each cluster to establish relationships between energy consumption and key explanatory variables, including environmental conditions and operational parameters. Compared to a single global model, the clustering–regression framework improves interpretability and allows for more accurate modelling within relatively homogeneous subsets of data. This approach is particularly suitable for high-performance buildings, where energy behavior is strongly influenced by operational schedules and control strategies rather than external conditions alone.

The regression models developed in this section are not intended as final predictive tools, but rather as an intermediate analytical step to support understanding of the building's energy behavior. Their primary objective is to quantify the relationship between energy consumption and key explanatory variables, including environmental conditions and operational parameters, within the context of the identified clusters. By doing so, the models provide insight into the relative importance of different input variables and help identify which parameters are most influential in predicting building energy use. These findings directly inform the selection of input variables and modelling assumptions for the parametric simulation framework developed in Chapter 5, thereby establishing a clear link between empirical data analysis and subsequent surrogate modelling.

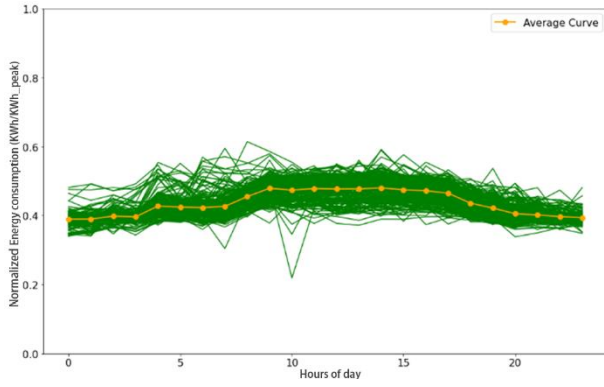
4.2.1 Clustering of the Total Building Energy Consumption Dataset

After data cleaning process and normalization, the k-means clustering method was applied to the total building energy consumption dataset, with the features for clustering being daily energy profiles. Every profile represents one day with 24 features, each representing the hourly energy consumption of the

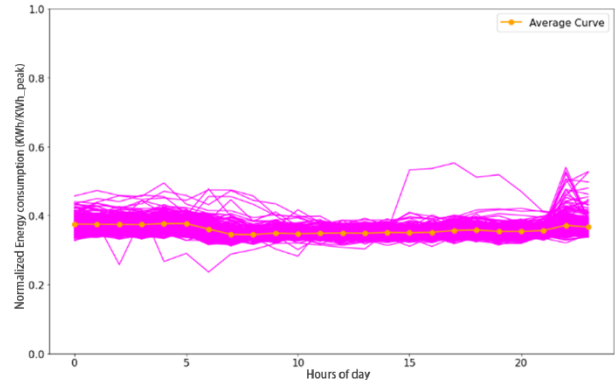
corresponding day (i.e. from 00:00 to 23:59). Using both the Silhouette index and elbow method, we identified seven as the optimal number of clusters for this dataset. K-means clustering was then applied to the total energy consumption data with $n=7$, as shown in Figure 4-4. Each resulting cluster reflects similar daily energy use patterns, enabling a more accurate analysis of variations in normalized daily consumption and offering deeper insights into the building's energy performance under different conditions.

The least daily energy consumption occurs under cluster-2 (magenta), characterized by relatively constant hourly usage. In contrast, cluster-3 (black) represents the highest daily energy consumption, typically associated with peak power demand periods. For most clusters, energy consumption starts to increase at 2:00 AM except for cluster-6 (green) and cluster-2 (magenta). This is because the HVAC schedule is programmed to exit setback mode at 2:00 AM and begin preheating or precooling to reach the setpoint by 7:00 AM. This control strategy will be discussed in more detail later. For cluster-6 (green), the energy consumption slightly increases around 7:00 AM, while cluster-2 (magenta) effectively presents a constant energy consumption throughout the day. To better understand the driving forces behind the seven daily energy profiles, a calendar map is developed, shown in Figure 4-5.

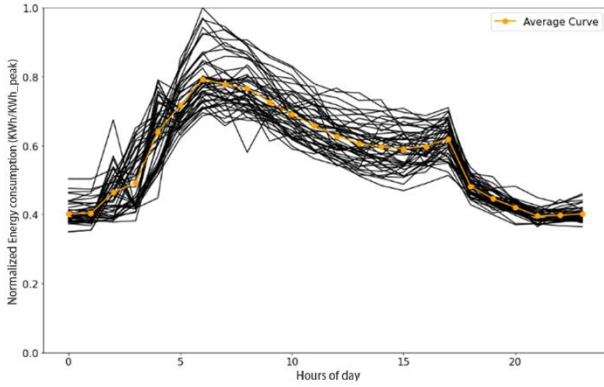
The calendar map shows that most clusters occur in specific periods of the year. For instance, the yellow cluster (Cluster-6) occurs only in winter 2021, while the blue cluster (Cluster-7) is limited to summer 2023. These distinct clustering patterns require further investigation to understand their underlying causes in terms of energy performance. Three distinct variable input parameters (i.e., features) were investigated for their influence on the building's energy consumption: i) occupancy levels, ii) HVAC control schedules and iii) weather conditions.



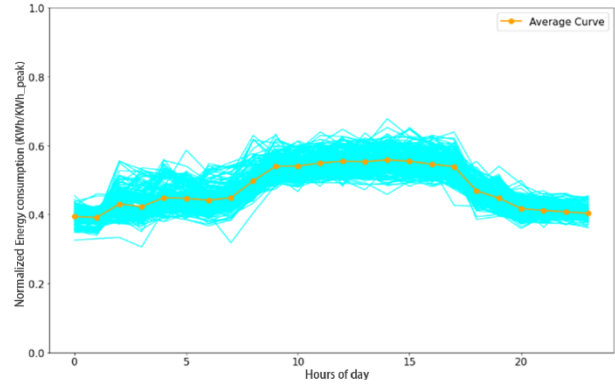
Total consumption_Cluster-1 (Green)



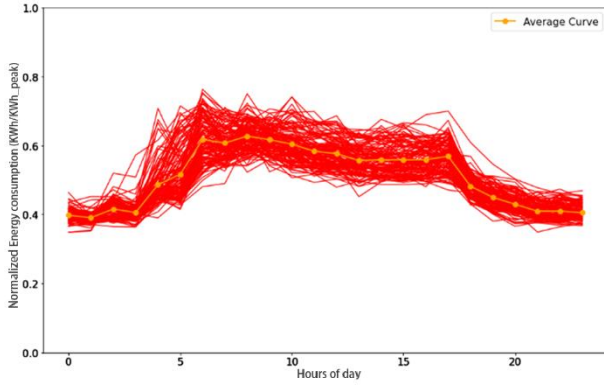
Total consumption_Cluster-2 (Magenta)



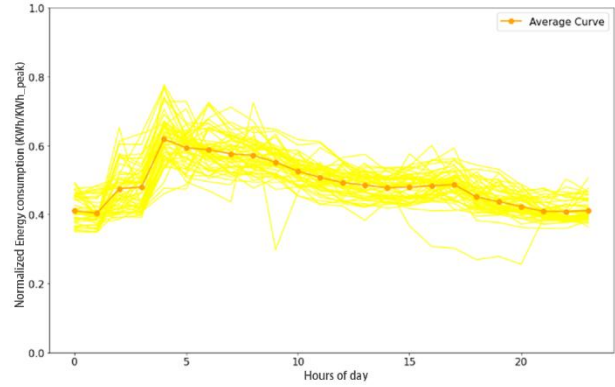
Total consumption_Cluster-3 (Black)



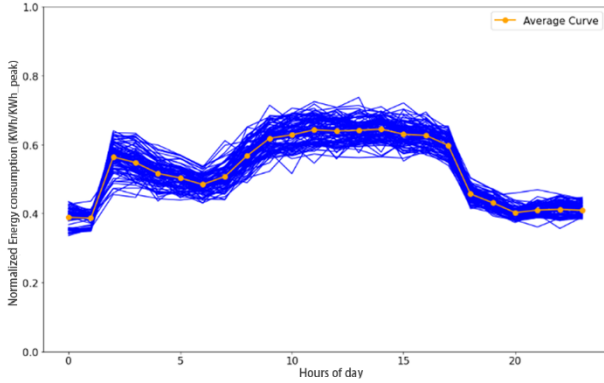
Total consumption_Cluster-4 (Cyan)



Total consumption_Cluster-5 (Red)



Total consumption_Cluster-6 (Yellow)



Total consumption_Cluster-7 (Blue)

Figure 4-4- Seven clusters for the normalized daily total energy consumption profiles using *k*-means clustering method. The orange dotted line represents the average daily profile for each cluster.

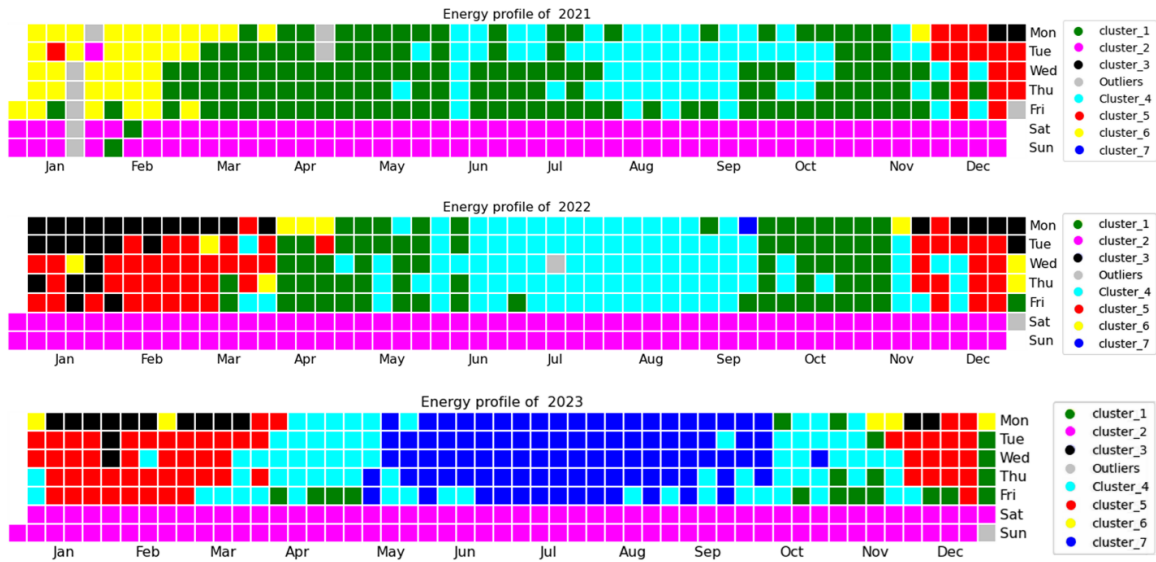


Figure 4-5- Calendar map for total daily energy profiles

Before the impact of these features is assessed, it is important to first break down total energy consumption into individual loads. The end-use energy breakdown presented in Figure 4-6 is derived from available submetered data within the building management system. Submeters were grouped into four primary categories, HVAC, lighting, plug loads, and domestic hot water (DHW), based on their functional classification, while remaining loads, including base building systems such as security, elevators, and water filtration, were categorized as “other loads.” This aggregation enables a simplified yet representative view of building energy consumption while maintaining consistency with the available measurement infrastructure. Figure 4-6 shows that HVAC systems account for the largest portion of energy consumption, while plug loads and lighting contribute comparable shares (23.5% and 20.5%, respectively), followed by smaller contributions from other loads.

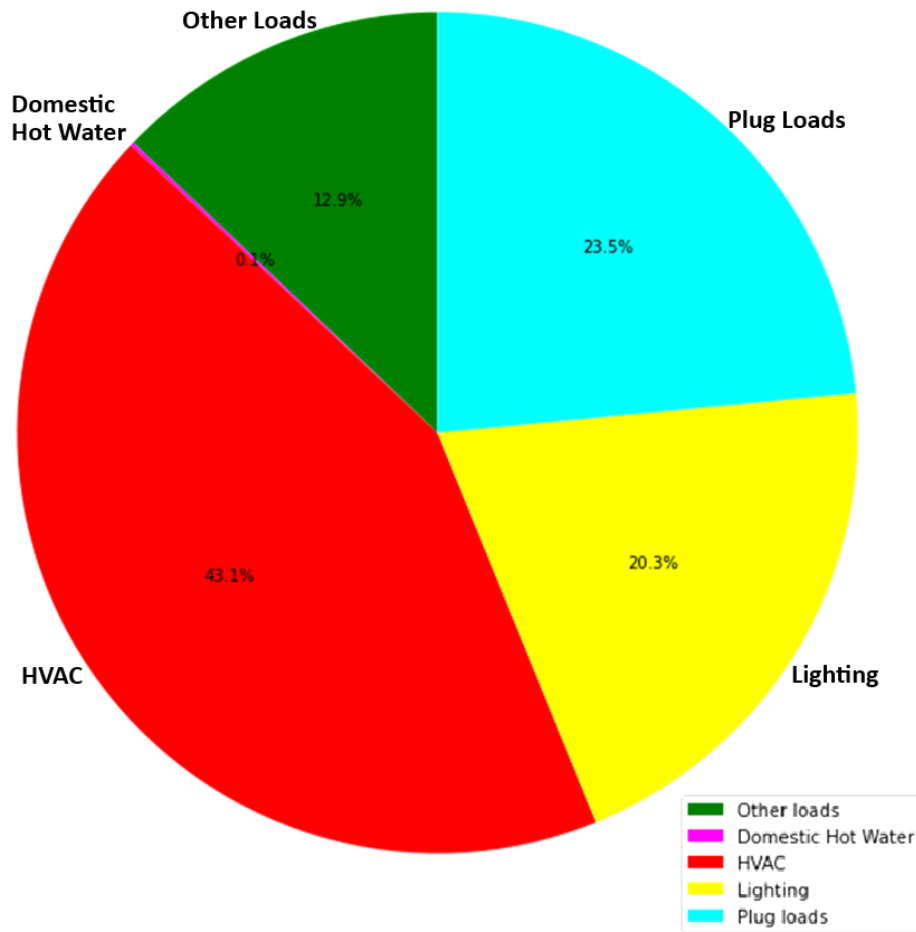


Figure 4-6- Percentage of individual loads over the total energy consumption of the Office Case Study.

4.2.2 Feature Analysis and Discussion of Individual Loads

The HVAC load analysis highlights the significant impact of HVAC control strategies on overall building energy consumption. To support this analysis, the occupancy schedule was derived through analysis of plug load data, which effectively captured the building's usage patterns. Although lighting load data was also examined, the results are not presented here, as they closely aligned with the occupancy trends identified from the plug load analysis and did not provide any additional insights.

Note that throughout this study, clusters are labeled using color-based identifiers (e.g., Red, Blue, Green) for ease of reference. These color labels are reused independently across different clustering tasks (e.g., total daily building energy profiles and individual load types such as HVAC or plug loads). Importantly, the same color does not imply any relationship or correspondence between clusters from different analyses. For example, the 'Red' cluster for HVAC loads is not necessarily related to the 'Red' cluster of total building energy profiles previously presented.

4.2.2.1 Impact of HVAC Controls

The energy profiles for HVAC energy consumption are presented in Figure 4-8, represented by six clusters. A calendar map was also developed (Figure 4-9). By inspecting the HVAC clusters and corresponding calendar map, we can see that cluster-HVAC3 represents the BMS setback for the weekends and holidays, resulting in nearly constant daily HVAC energy consumption. This indicates that the cluster is not season-dependent, suggesting that the building setpoints might have a greater influence on energy consumption than weather conditions. Additionally, cluster-HVAC5 represents HVAC operation for the cooling season during 2021 and 2022, while cluster-HVAC2 represents the 2023 HVAC consumption during cooling season. Cluster-HVAC1 corresponds to the shoulder seasons (i.e., spring and fall) for 2021 and 2022, while cluster-HVAC5 represents the similar time periods for 2023. Cluster-HVAC4 and cluster-HVAC6 represent the heating season, with heating peaks occurring in Cluster-HVAC4. Notably, except for cluster-HVAC3, which represents weekends and holidays, all other clusters show an energy consumption ramp-up starting at 2:00 AM and peaking between 5:00 AM and 9:00 AM, as expected. As previously mentioned, this is because the BMS is programmed to exit setback mode at 2:00 AM and begin preheating or precooling to reach the setpoint by 7:00 AM, minimizing peak demand charges. All clusters exhibit a common pattern of reduced consumption during the late afternoon and evening.

Based on the consistent weekday patterns observed across all HVAC energy clusters and the corresponding temperature setpoint schedules, a representative normalized daily HVAC operational schedule was developed for typical workdays. As illustrated in Figure 4-10, the proposed schedule captures the preheating/precooling strategy initiated at 2:00 AM, which shifts and distributes energy consumption across the early morning hours to effectively mitigate the morning peak demand. This operational strategy reduces peak electrical demand and alleviates stress on the utility grid, as also supported in the literature [130]. The same normalized weekday pattern is evident across all HVAC clusters representing occupied days. For weekends and holidays, a normalized constant HVAC schedule with a value of 0.2 was adopted to reflect the observed setback operation. The HVAC schedule shown in Figure 4-10 is subsequently evaluated for its effectiveness as an input variable in the building energy consumption prediction models.

The HVAC schedule is defined as a normalized control signal ranging between 0 and 1, where 0 represents setback operation and 1 corresponds to full system operation. The schedule is derived from the HVAC setpoint temperature profiles and represents the average behavior across the three building floors, each of which operates with slightly different setpoint schedules. Although the schedule conceptually reflects a gradual ramp-up of system activity during transition periods (e.g., early morning

preconditioning), it is implemented as a step function, as shown in Figure 4-10, where discrete changes approximate the operational behavior of the building management system.

4.2.2.1.1 An Example of Fault Detection Diagnostics Through Clustering

Although fault detection is outside the primary scope of this thesis, this finding demonstrates an additional application of the clustering methodology and validates its sensitivity to operational anomalies. In 2023, a distinct deviation in the HVAC energy consumption pattern was detected through clustering results (Cluster-HVAC2), indicating a potential operational fault and motivating a targeted inspection of the building's mechanical system. It was revealed that the water loop temperature, connecting the geothermal loop to both the VRF system and a three-way heat pump, was incorrectly set to lower temperature during the spring and summer of 2023, leading to an increase in energy consumption. As a result, Cluster-7 (blue, in Figure 4-4) was considered invalid for daily energy profile prediction and was excluded from further analysis, with the focus placed on the remaining valid clusters. This analysis highlights that the clustering approach can also serve as a useful tool for fault detection in building systems, helping to identify operational anomalies efficiently.

To verify whether restoring the water loop temperature to its previous value resolves the issue of anomalous clustering, a focused analysis was conducted using the total energy consumption data for 2024. The objective was to demonstrate that correcting the water loop fault setpoint would restore the clustering behavior to its pre-fault state. As shown in Figure 4-7, the calendar map for 2024 indicates that Cluster-7 (blue) has nearly disappeared. The few remaining scattered blue points likely represent borderline profiles that, in the absence of the dominant blue cluster observed in 2023, would have merged with adjacent clusters.

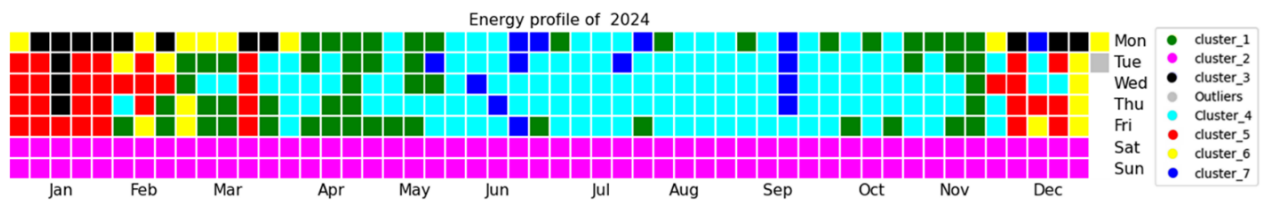


Figure 4-7- Calendar map for total energy consumption in 2024

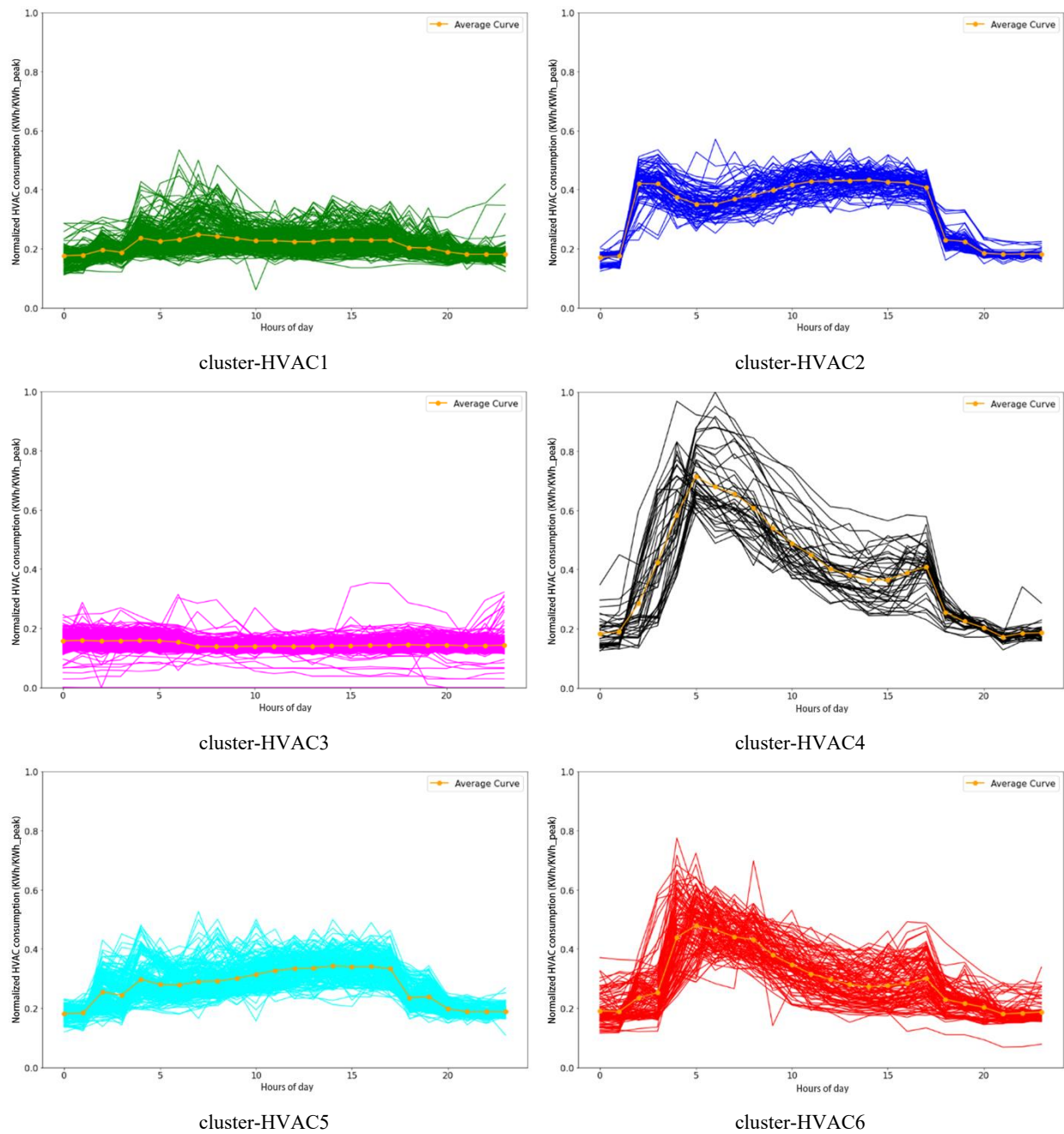


Figure 4-8- HVAC energy clusters using *k*-means method.

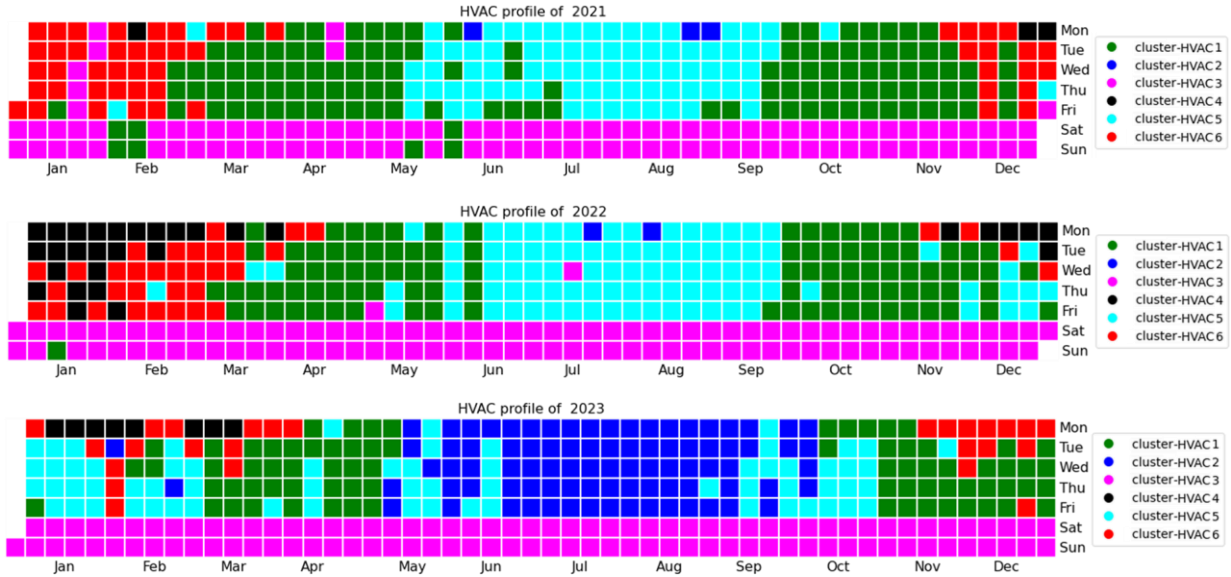


Figure 4-9-Calendar map for HVAC energy consumption profiles

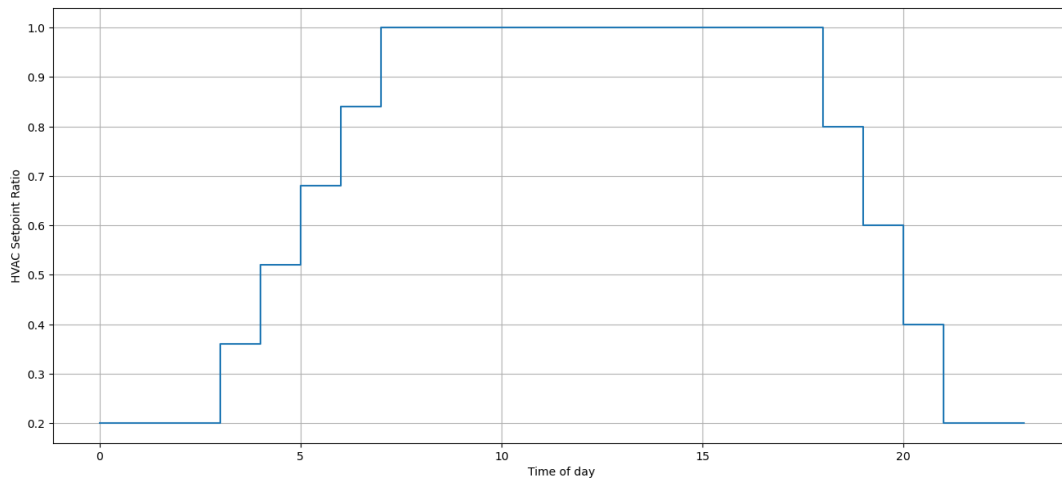


Figure 4-10. Proposed normalized daily HVAC schedule during workdays

4.2.2.2 Impact of Plug Loads and Occupancy Levels

While HVAC energy use is primarily governed by control strategies and weather conditions, plug load consumption is directly driven by occupant presence and behavior. In this study, occupancy patterns were substantially disrupted by the COVID-19 pandemic, during which Canadian office buildings remained largely unoccupied between January and September 2021 as most employees transitioned to remote work. As a result, many building owners adjusted setpoint temperatures and schedules to reduce the energy consumption of unoccupied office buildings. Occupancy-related data, such as CO₂ concentration levels in exhaust air and occupancy sensor data, can be employed to assess the impact of occupancy on building energy consumption. Alternatively, in the absence of direct occupancy data, an alternative approach involves analyzing plug loads which are directly linked to building occupancy. The latter method

provides the possibility of indirect estimation of occupancy levels and contributes to a better understanding of how occupancy can be used to predict the daily energy profiles of a building.

Initially, four clusters were detected for plug loads consumption data, presented in Figure 4-11. Cluster-PL2 and cluster-PL4 exhibit a similar plug load consumption pattern, characterized by a relatively constant consumption trend, with cluster-PL2 having a higher plug load energy consumption compared to cluster-PL4. These two clusters evidently correspond to periods of either vacancy or very low occupancy, with subtle difference in the base loads of the building. On the other hand, cluster-PL1 and cluster-PL3 share a similar consumption pattern. It is clear that both clusters represent plug load consumption driven by a surge in occupancy levels in the morning and a gradual reduction in occupancy during late afternoon.

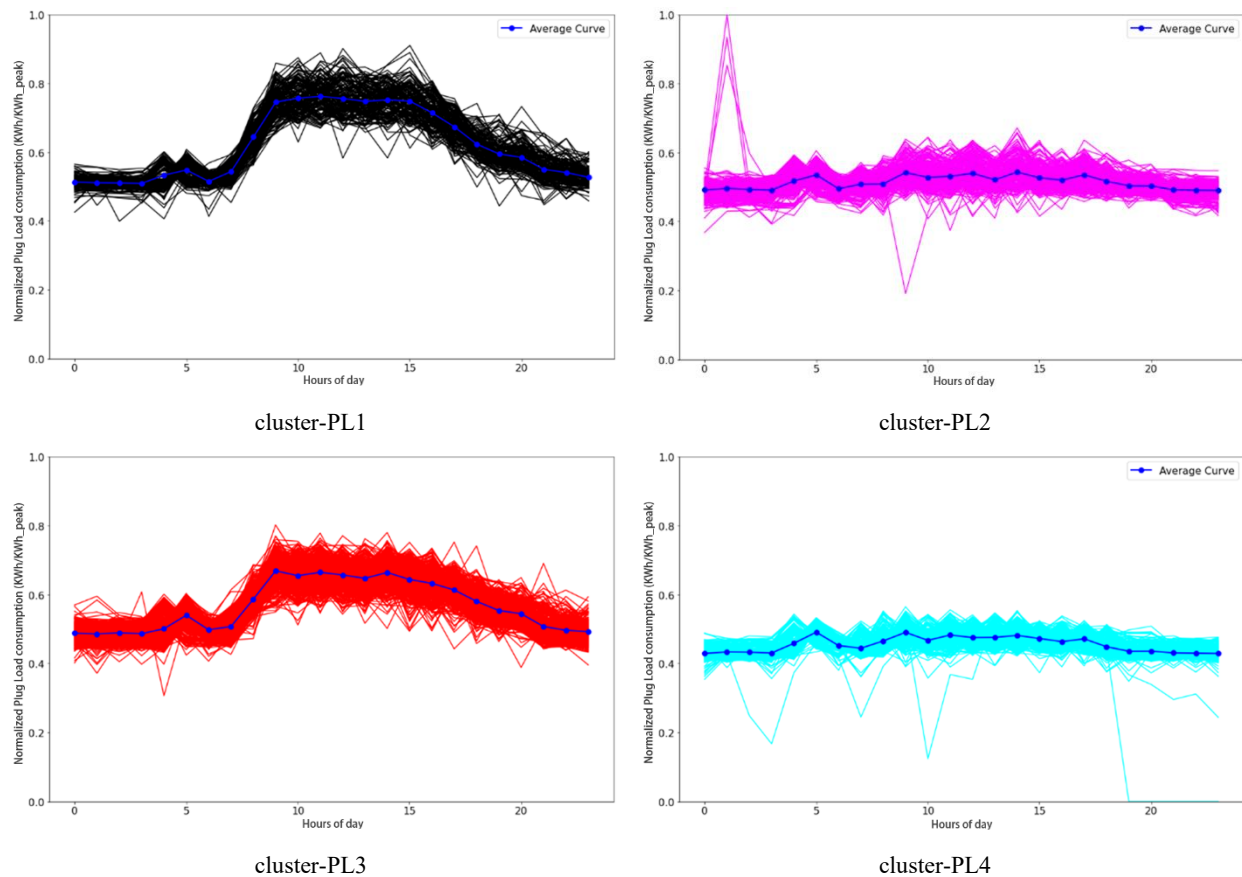


Figure 4-11- Plug loads clusters using k-means method

Figure 4-12 presents a calendar map showing the temporal distribution of each cluster. The color map and plug load profiles reveal four distinct periods: two with low occupancy – one during the pandemic (cluster-PL4) and the other post-pandemic on weekends and holidays (cluster-PL2) – and two with increased occupancy. The first period of increased occupancy occurs at the end of the pandemic (late 2021) as people gradually returned to the office (cluster-PL3) and a somewhat normal occupancy pattern

started to show based on the plug load consumption. It should be mentioned that a temporary rise in occupancy (cluster-PL1) from fall 2022 till May 2023 was detected, due to a few short-term tenants.

Based on the similar pattern shared by cluster-PL1 and cluster-PL3, a normalized daily schedule for increased occupancy levels was developed. It features a sharp rise in occupancy in the morning, followed by a gradual decline in the evening. The occupancy schedule is illustrated in Figure 4-13. For holidays, weekends or temporary office closure days (e.g., during pandemic), a normalized schedule with a constant value of 0.0 was used to represent building vacancy during these periods.

A summary description of the plug load profiles and their relation to occupancy is presented in Table 4-2, providing insight into selecting input parameters for predicting the building's daily energy consumption.



Figure 4-12- Calendar map for plug loads profiles

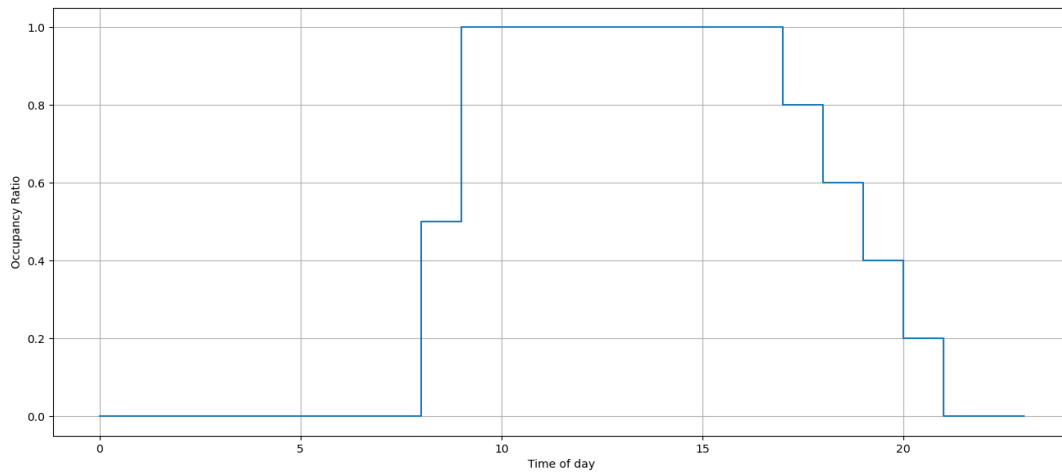


Figure 4-13- Representative schedule for periods of increased occupancy, generated based on plug load cluster-PL1 and cluster-PL3.

Table 4-2- Plug loads characteristics and its relation to occupancy patterns

Cluster #	Plug loads profile characteristics	Occupancy pattern
PL1	Variable increased plug loads consumption	Post-pandemic; Weekdays (temporary rise of occupancy due to temporary lease contract)
PL2	Nearly constant reduced plug load consumption	Post-pandemic; weekends and holidays or limited return-back-to-telework periods
PL3	Variable increased plug loads consumption	Transition from pandemic to post-pandemic Weekdays (normal occupancy)
PL4	Nearly constant reduced plug load consumption	During pandemic Weekdays (telework)

4.2.2.3 Predicting the Daily Energy Profile of the Office Case Study

4.2.2.3.1 Weather Conditions as Energy Predictors

Weather conditions tend to be a dominant input parameter in the building energy prediction models. For this Office Case Study, the exterior temperature was recorded and used as one of the inputs to predict energy consumption. However, solar irradiance is not measured directly by BMS. Instead, the PV electricity generation was used as an equivalent “solar irradiance” input as it is linearly related to the total horizontal solar irradiance. As literature indicates, most data-driven energy models have used outside temperature and solar irradiance as inputs to their models to predict building energy consumption [24], [132], [133]. In this study for each cluster, exterior temperature and irradiance (i.e., in the form of PV electricity generation) were used as input(s) in i) polynomial regression model, and ii) MLP artificial neural network. Table 4-3 displays the prediction outcomes for these models.

The hyperparameters of the MLP model were tuned with Optuna using the Tree-structured Parzen Estimator Sampler (TPESampler) [134]. The tuning process was carried out over 250 trials, exploring a range of configurations including hidden layer sizes [i.e., (50,), (100,), (1000,), (50,50), (100,50)], activation functions (i.e., *relu*, *tanh*, *logistic*), solver (i.e., *adam*, *lbfgs*), and learning parameters such as regularization term (*alpha*) and Initial learning rate, both sampled on a logarithmic scale. The model was trained using 5-fold cross-validation to maximize the R^2 score, and the final model was evaluated on a test set using the best-performing hyperparameters.

Interestingly, even though the RMSE is not high (i.e. less than 4 Wh/m²), there is a very low correlation between weather parameters and the total energy consumption of the building for most of the clusters. Furthermore, MAPE is relatively high, reducing the quality of the fitting results. Besides cluster-4 (blue) that has a somewhat strong correlation to radiation input parameter, for all other clusters, the dependency to temperature and/or radiation is considerably low. This is because the building is not conventional but rather a high-performance one. Its energy performance is less sensitive to outdoor conditions due to high

efficiency mechanical system that includes a geothermal heat pump, a VRF system, an energy recovery system (ERV) and an enhanced building envelope, which provides high insulation levels, airtightness, and moderate solar heat gains. Additionally, the primary heating and cooling source for the building is an open-loop ground source heat pump, which benefits from the nearly constant underground water temperature throughout the seasons. This observation is counterintuitive to the common belief that outdoor conditions primarily drive building energy consumption. While this holds true for conventional buildings, this study provides evidence that it may not apply to high-performance buildings. The findings indicate that outdoor temperature and solar radiation cannot be used alone as reliable predictors to tell when a cluster would occur.

Table 4-3- Prediction results for all the six clusters, with weather parameters as input(s).

Polynomial regression							
Input parameters	Metric	Cluster-1* (green)	Cluster-2* (magenta)	Cluster-3* (Black)	Cluster-4* (Cyan)	Cluster-5* (red)	Cluster-6* (yellow)
Outdoor temperature	RMSE (kWh/m ²)	1.04e ⁻³	4.69e ⁻⁴	3.83e ⁻³	1.55e ⁻³	2.38e ⁻³	1.90e ⁻³
	R ²	0.05	0.13	0.025	0.1	0.009	0.05
	MAPE	10.95%	6.15%	25.45%	16.24%	23.34%	17.16%
Radiation	RMSE (kWh/m ²)	8.6e ⁻⁴	4.68e ⁻⁴	3.58e ⁻³	1.11e ⁻³	2.29e ⁻³	1.97e ⁻³
	R ²	0.32	0.14	0.057	0.56	0.092	0.02
	MAPE	8.58%	5.98%	24.60%	8.96%	19.83%	17.62%
Outside temperature + Radiation	RMSE (kWh/m ²)	7.9e ⁻⁴	4.44e ⁻⁴	3.45e ⁻³	1.06e ⁻³	2.24e ⁻³	1.63e ⁻³
	R ²	0.34	0.24	0.1	0.59	0.16	0.091
	MAPE	7.68%	5.69%	22.44%	7.9%	16.67%	14.07%
Multi-Layer Perceptron (MLP)							
Outdoor temperature	RMSE (kWh/m ²)	1.04e ⁻³	4.95e ⁻⁴	3.7e ⁻³	1.53e ⁻³	2.27e ⁻³	1.82e ⁻³
	R ²	00.06	0.06	0.04	0.11	0.009	0.076
	MAPE	10.52%	6.75%	25.24%	15.86%	22.88%	16.80%
Radiation	RMSE (kWh/m ²)	8.4e ⁻⁴	4.42e ⁻⁴	3.35e ⁻³	1.03e ⁻³	0.076	1.96e ⁻³
	R ²	0.38	0.24	0.16	0.57	0.303	0.03
	MAPE	6.22%	5.53	19.8%	8.7%	11.95%	17.51%
Outside temperature + Radiation	RMSE (kWh/m ²)	8.54e ⁻⁴	4.4e ⁻⁴	3.2e ⁻³	9.87e ⁻⁴	1.87e ⁻³	1.60e ⁻³
	R ²	0.37	0.36	0.19	0.60	0.312	0.13
	MAPE	6.4%	4.63	19.62%	7.6%	11.75%	12.8%

*The six clusters correspond to total energy consumption as presented in Figure 4-4 (Note: Cluster-7 was identified as a faulty cluster and was excluded)

4.2.2.3.2 HVAC and Occupancy as Additional Model Predictors

For better prediction results additional parameter inputs are needed. The following regression analysis accounted for weather conditions (i.e., outdoor temperature and PV generation as a “solar radiation” feature), occupancy levels and HVAC schedules. The analysis was performed for various input combinations. Although a wide range of machine learning and regression techniques could be applied to this problem, this study focuses exclusively on polynomial regression and MLP models at this stage. This deliberate restriction allows for a controlled comparison between two fundamentally different modelling paradigms, an interpretable parametric approach and a flexible nonparametric learning model, while

prioritizing the development of a data-driven prediction framework that can be readily implemented within conventional building management systems. Importantly, both models can be deployed without requiring additional computational resources or reliance on cloud-based services. Table 4-4 outlines the minimum set of predictors required for each cluster to accurately model total energy consumption using polynomial regression and MLP. Note that these six clusters correspond to distinct patterns in energy usage as depicted in Figure 4-4 except for blue cluster (i.e., Cluster-7) which has been considered as faulty behavior of HVAC system and was excluded from Table 4-4. The model's R^2 values exceeded 0.80 for several clusters, confirming the value of incorporating these internal drivers. This finding aligns with the broader understanding that, as buildings become more energy-efficient, their performance becomes increasingly sensitive to occupant behavior and system control strategies rather than external weather variables alone.

Table 4-4- Input parameters and fitting results for polynomial regression and MLP on the total energy consumption of the building.

		Cluster-1* (green)	Cluster-2* (magenta)	Cluster-3* (black)	Cluster-4* (cyan)	Cluster-5* (red)	Cluster-6* (yellow)
Input Parameters	Metric	1-Outside Temperature 2- Radiation 3-HVAC Schedule	1-Outside Temperature 2- Radiation	1-Outside Temperature 2- Radiation 3-HVAC Schedule 4-Occupancy Schedule	1-Outside Temperature 2- Radiation 3-HVAC Schedule 4-Occupancy Schedule	1-Outside Temperature 2- Radiation 3-HVAC Schedule 4-Occupancy Schedule	1-Outside Temperature 2- Radiation 3-HVAC Schedule
Polynomial Regression Fitness	RMSE (kWh/m ²) R^2 MAPE	6.66e ⁻⁴ 0.69 4.70%	4.44e ⁻⁴ 0.25 5.69%	1.43e ⁻³ 0.86 7.74%	8.37e ⁻⁴ 0.83 4.68%	1.06e ⁻³ 0.82 6.20%	1.42e ⁻³ 0.41 9.04%
MLP Regression Fitness	RMSE (kWh/m ²) R^2 MAPE	6.66e ⁻⁴ 0.73 4.44%	4.4e ⁻⁴ 0.36 4.63	1.39e ⁻³ 0.88 7.59%	8.31e ⁻⁴ 0.84 4.37%	1.01e ⁻³ 0.83 6.09%	1.23e ⁻³ 0.47 8.68%

*The six clusters correspond to total energy consumption as presented in Figure 4-4 (Note: Cluster-7 was identified as a faulty cluster and was excluded).

For Cluster-2 (magenta), which represents weekends and holidays, occupancy and HVAC schedule cannot be used as input parameters, this is because both the occupancy and HVAC setpoint schedules are assigned constant values for weekends and holidays, resulting in a single fixed value that cannot be used as a dynamic input parameter. Similarly, for Cluster-6 (yellow), which corresponds to the pandemic period, occupancy was also excluded as an input. For both clusters, temperature and “solar radiation” were selected as inputs. For other clusters, temperature, occupancy, and HVAC schedule were used as input parameters. A second-order polynomial was selected as providing the best balance between predictive performance and interpretability for BMS deployment, with cross-product terms retained to capture interactions between environmental variables and operational states.

Based on the identified relevant inputs for each cluster, a generalized regression structure was formulated to capture both individual effects and interactions between variables. The selected formulation balances model interpretability with sufficient flexibility to represent nonlinear relationships observed in the data.

As shown in Table 4-4, outdoor temperature and solar radiation alone yield $R^2 \leq 0.41$ for the primary workday clusters; the dominant predictors are HVAC schedule and occupancy, variables that change-point models do not accommodate. When these operational variables are included, model performance improves significantly ($R^2 \approx 0.82\text{--}0.86$), demonstrating that energy consumption in this high-performance building is not governed by temperature alone. Furthermore, the building operates in distinct clusters with different predictor relationships, rather than following a single continuous temperature-dependent response.

The general form of the second-order polynomial regression equation with four input parameters is formulated below, with the additional constraint that predicted energy values remain non-negative:

$$E_{Norm} = \max(f(T, P, S, O), 0) \quad \text{Eq 4-1}$$

where

$$\begin{aligned} f(T, P, S, O) = & a_0 + a_T T + a_P P + a_S S + a_O O + a_{TT} T^2 + a_{PP} P^2 + a_{SS} S^2 + a_{OO} O^2 \dots \\ & + a_{TP} T \cdot P + a_{TS} T \cdot S + a_{TO} T \cdot O + a_{PS} P \cdot S + a_{PO} P \cdot O + a_{SO} S \cdot O \end{aligned}$$

where E_{Norm} is the normalized total building electricity consumption; T is the outside temperature in $^{\circ}\text{C}$; P is the normalized PV generation; S is the HVAC set point schedule (0 to 1); and, O is the Occupancy schedule (0 to 1). Then, the actual energy consumption of the Office building at any given moment can be estimated as follows:

$$E_{total} = E_{Norm} \cdot E_{peak} \text{ (kWh)} \quad \text{Eq 4-2}$$

where E_{peak} the hourly peak electricity consumption (kWh) of the building (i.e., $E_{peak} = 232 \text{ kWh}$).

The identified clusters were organized into a hierarchical tree structure to provide a clear and logical interpretation of the daily energy profile groups derived from the clustering process. This structure highlights the key features that differentiate operational regimes and supports the prediction of daily energy behavior based on a limited set of input variables. Figure 4-14 presents this tree-based representation, illustrating how distinct patterns in energy consumption can be systematically categorized. This representation provides an intuitive overview of building operation and serves as a bridge between detailed time-series data and simplified predictive modelling.

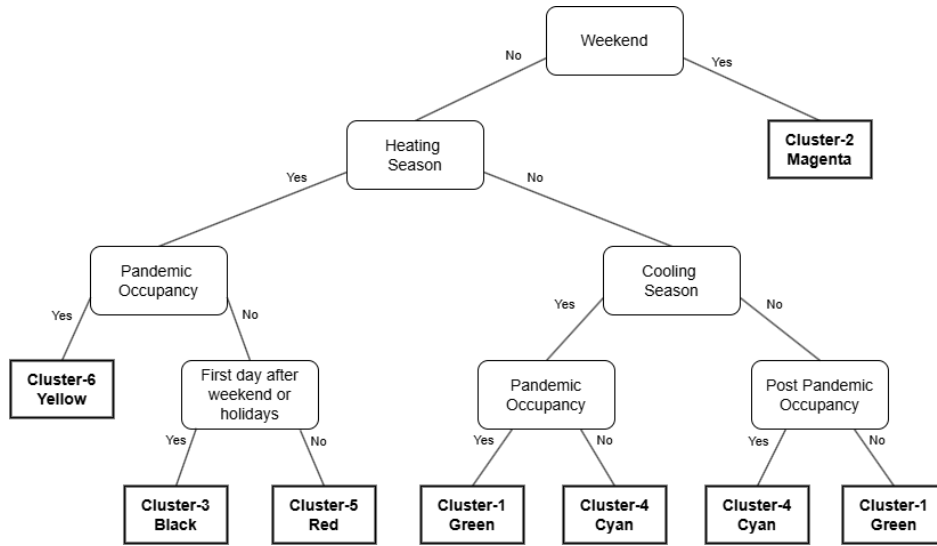


Figure 4-14- A tree flowchart for total energy consumption clusters as appear in Figure 4-8.

Note that univariate change-point models such as ASHRAE Guideline 14 3P/5P CPM [135] were considered but found structurally insufficient, as they assume outdoor temperature as the sole driver, a condition the regression results above confirm does not hold for this building. Multivariate extensions exist but would require structural modifications that move beyond the standard CPM framework, at which point a polynomial regression formulation offers equivalent flexibility with greater interpretability for BMS deployment.

4.3 Remarks on the Case Study Analysis

The case study analysis presented in this chapter provides key insights into the energy behavior of high-performance office buildings and establishes the empirical foundation for the subsequent phases of this research. Through clustering and regression-based analysis of high-resolution operational data, several important observations can be made.

First, internal drivers, namely occupancy schedules, HVAC control strategies, and internal loads, play a more significant role in shaping hourly energy consumption than external environmental variables. Regression models using weather inputs alone yielded $R^2 \leq 0.41$ for workday clusters, while incorporating HVAC schedules and occupancy improved performance to $R^2 = 0.82\text{--}0.88$, quantitatively confirming that energy consumption in this high-performance building is not governed by outdoor conditions alone, a finding that is counterintuitive relative to conventional buildings but expected for a high-performance building where improved envelope performance, reduced infiltration, high-efficiency HVAC systems, and ground-source heat pump operation collectively reduce sensitivity to outdoor conditions.

Second, the application of clustering techniques proved effective in simplifying the interpretation of complex time-series data. Combined with calendar-based visualization, this representation supports identification of seasonal trends, peak demand periods, and operational anomalies, including the water loop fault detected in 2023, without requiring direct inspection of raw time series. This clustered representation is carried forward in Chapter 6 as an interpretive tool for analyzing surrogate model outputs.

Chapter 5

5 Parametric Numerical Building Energy Model: Development and Verification

Building on the empirical findings of Chapter 4, this chapter presents the development, parametrization and verification of a custom EnergyPlus building energy model to verify its numerical correctness and physical consistency prior to its use in generating synthetic training data for surrogate model development. The analysis is conducted under the foundational assumption that EnergyPlus simulations [8], [119], [122] can reliably represent real building performance, an assumption supported by extensive validation studies reported in the literature.

The credibility of EnergyPlus is supported by extensive testing against standardized validation procedures (e.g., ANSI/ASHRAE Standard 140/BESTEST [136]) and by multiple empirical studies comparing EnergyPlus predictions with high-quality measured data from instrumented test facilities and real buildings [120], [137], [138]

Given that the parametric model has been specifically developed for this research, it is necessary to perform formal verification to ensure that the implementation is free of formulation or coding errors and that its responses remain plausible across the full parameter space. Because the parametric model has been custom-built for this research rather than adopted from an established benchmark, formal verification is conducted to confirm that all parameterizations produce physically consistent responses across the full input space.

To accomplish this, a series of hypotheses is evaluated using a randomly generated building geometry while systematically varying all design parameters across their allowable ranges. The objective is to assess whether model outputs exhibit trends consistent with established physical principles, building physics theory, and known directional sensitivities. This process functions as a quality-assurance stage prior to data generation, ensuring that any synthetic data used for surrogate model training reflects valid physical relationships rather than artifacts of model formulation. Without this verification step, systematic modelling errors could propagate into the surrogate model and compromise its predictive reliability.

5.1 Model Description

This research relies on a comprehensive parametric physics-based EnergyPlus (v25.1) model as the foundational building representation for generating the synthetic dataset used in training and evaluating the surrogate model. Before developing the surrogate framework, the parametric detailed model was

carefully constructed and reviewed to ensure that it can be used for synthetic data generation, with reliable thermal performance, HVAC operation, and indoor environmental quality.

In this section model architecture and parameters of the numerical model which are not changing through data generation process would be explained to have better understanding of crucial design of the parametric numerical model.

5.1.1 Building Geometry and Thermal Zoning

The parametric model represents a low and midrise commercial office buildings located in Waterloo, Ontario, Canada. The geometry follows ASHRAE zoning recommendations with a rectangular footprint configured as one core zone and four perimeter zones per floor, with a perimeter zones always be 5-meter deep.

This zoning strategy is able to capture the impact of solar and internal gains, and HVAC controls while ensuring consistent parameter modifications (aspect ratio, WWR, envelope thermal resistance, orientation) across all floors.

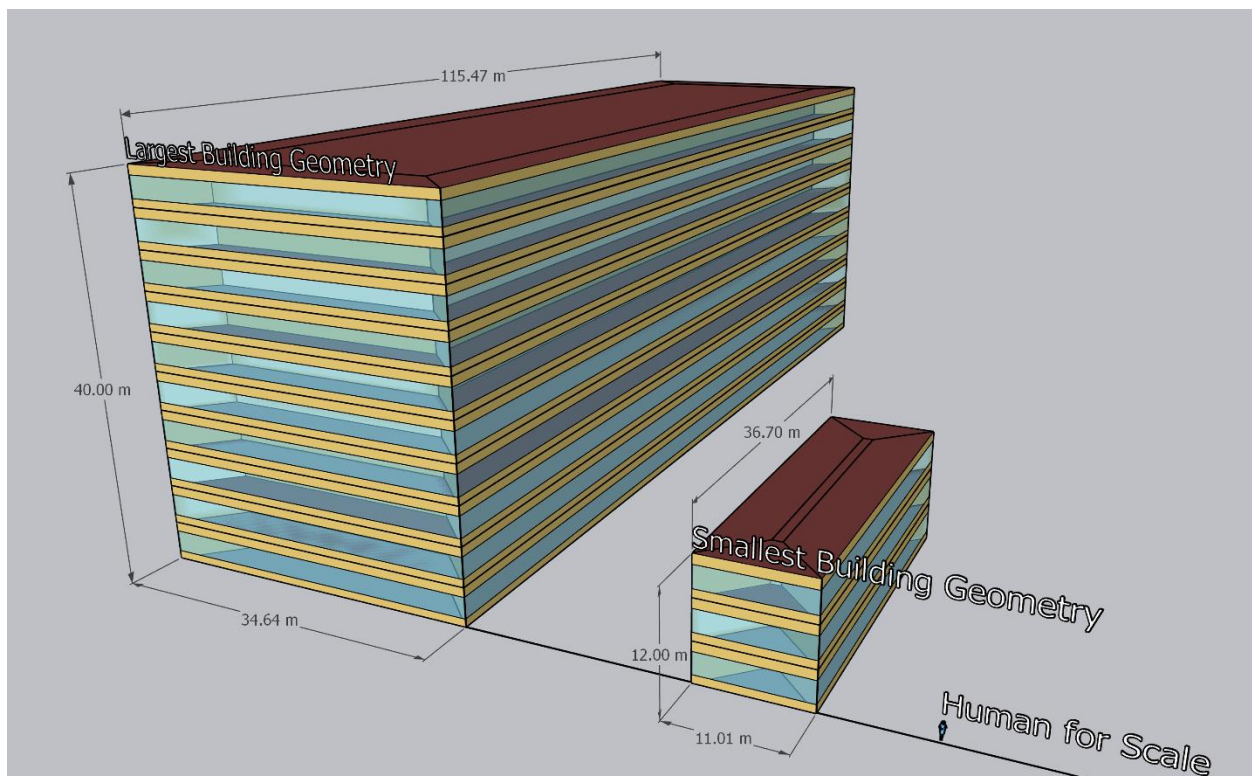


Figure 5-1- Schematic model geometry and zoning for smallest and largest office buildings represented by the parametric model

5.1.2 Envelope Characteristics

The building envelope is modeled in accordance with ASHRAE 90.1 and NECB requirements. While the overall construction typologies are fixed to represent a typical high-performance office building, several envelope-related performance properties are treated as parametric variables within predefined ranges, as described in the Parameter Range selection section. To balance building physics with parametric flexibility, the envelope components are handled as follows:

Wall and Roof Assemblies: Standard office construction assemblies were selected as baseline configurations. The thermal performance of these assemblies was parameterized by varying the insulation thickness, allowing wall and roof thermal resistance to be treated as continuous parametric inputs while maintaining realistic layer compositions.

Ground Floor: The ground floor is modeled as a slab-on-grade construction with effective RSI-3.9 similar to the case study building.

Windows: Double-glazed window units were assumed. Within fixed glazing assembly, window U-factor and SHGC were parameterized to represent a range of commercially available high-performance window products.

5.1.3 Internal Loads and Occupancy Profiles

Internal gains follow ASHRAE 90.1 standard with schedules used in National Energy Code of Canada for Buildings for office buildings (NECB) 2020, presented in Figure 5-2. Occupancy Characteristics are listed as:

- **Occupancy Peak Density:** 0.056 person/m² in office areas
- **Heat Gains due to Occupancy:** 75 W/person (sensible), 45 W/person (latent)
- **Occupancy Schedule:** based on NECB (Figure 5-2)
- **Peak Occupancy Hours:** 9 AM - 4 PM with 90% maximum occupancy

Lighting and Equipment Power Density are considered as parametric inputs and their range has been presented in the parameter range selection section. However, standard schedule for lighting and Equipment has been adopted from NECB which have been presented in the Figure 5-2. The values for these charts have been presented in the **Error! Reference source not found..**

5.1.4 HVAC System Configuration and Performance

For this study, a single mechanical system has been considered, modeled after the high-performance office case study building (Chapter 4). Heating and cooling for the office zones was provided by a Variable Refrigerant Flow (VRF) system. Additionally, a rooftop packaged unit equipped with an enthalpy wheel supplies conditioned fresh air and facilitate heat recovery. To meet any remaining heating and cooling demands after heat recovery, supplementary coils are included within the fresh air system.

The choice of using only a VFR system is justified on two grounds. First, heat-recovery VRF systems are among the most energy-efficient all-electric HVAC technologies available for commercial buildings, and are explicitly identified in the literature as a viable pathway to achieving net-zero and nearly zero-energy building performance [139]. Second, VRF heat recovery systems are uniquely capable of simultaneously transferring heat between zones with complementary loads, such as interior server rooms requiring year-round cooling and perimeter zones requiring heating, without generating thermal energy redundantly. This load-shifting capability is a performance requirement in modern office buildings with mixed-use zones and represents a key reason the system is prevalent in high-performance commercial construction.

Operational schedules are held fixed throughout parametric simulations rather than treated as variable inputs. Occupancy, lighting, and equipment schedules follow standard profiles derived from ASHRAE 90.1 [9] and NECB 2020 [10], consistent with established practice in parametric building performance simulation where standard schedules are used to isolate the influence of design parameters under representative operational assumptions. HVAC setpoint schedules, by contrast, are derived from the operational pattern of the case study building analyzed in Chapter 4, which demonstrated that the pre-conditioning strategy, initiating HVAC ramp-up at 2:00 AM to reach setpoint by 7:00 AM, is a significant driver of time-resolved energy behavior. These schedules are presented in Figure 5-3 and tabulated in

Appendix Table 2. Note incorporating operational schedules as parametric inputs would require vector-based representations of full-year hourly profiles, while also introducing categorical treatment of multiple schedule types, both of which are identified as directions for future research.

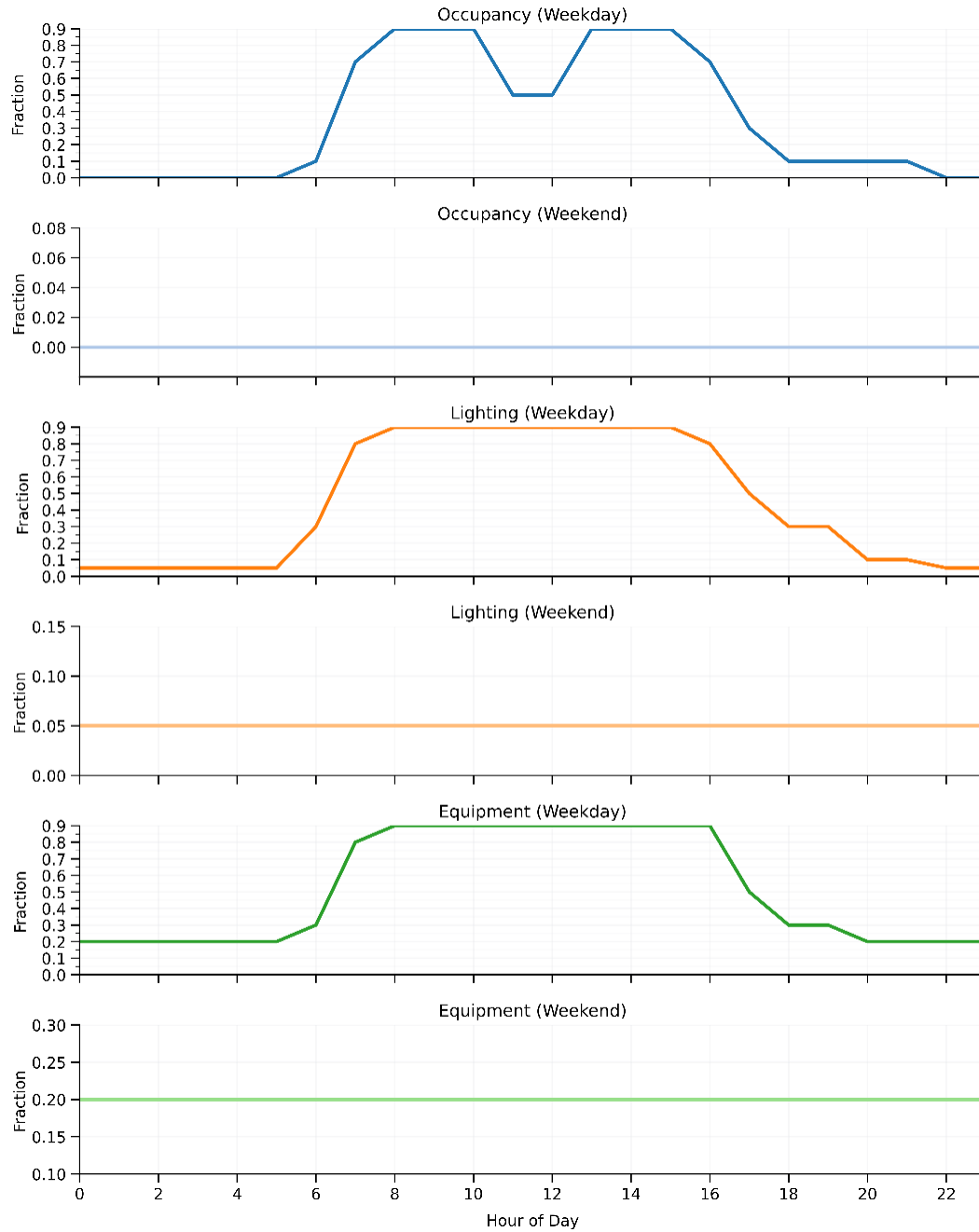


Figure 5-2- Occupancy, Lighting and Equipment schedule for office buildings based on NECB (adopted from [10])

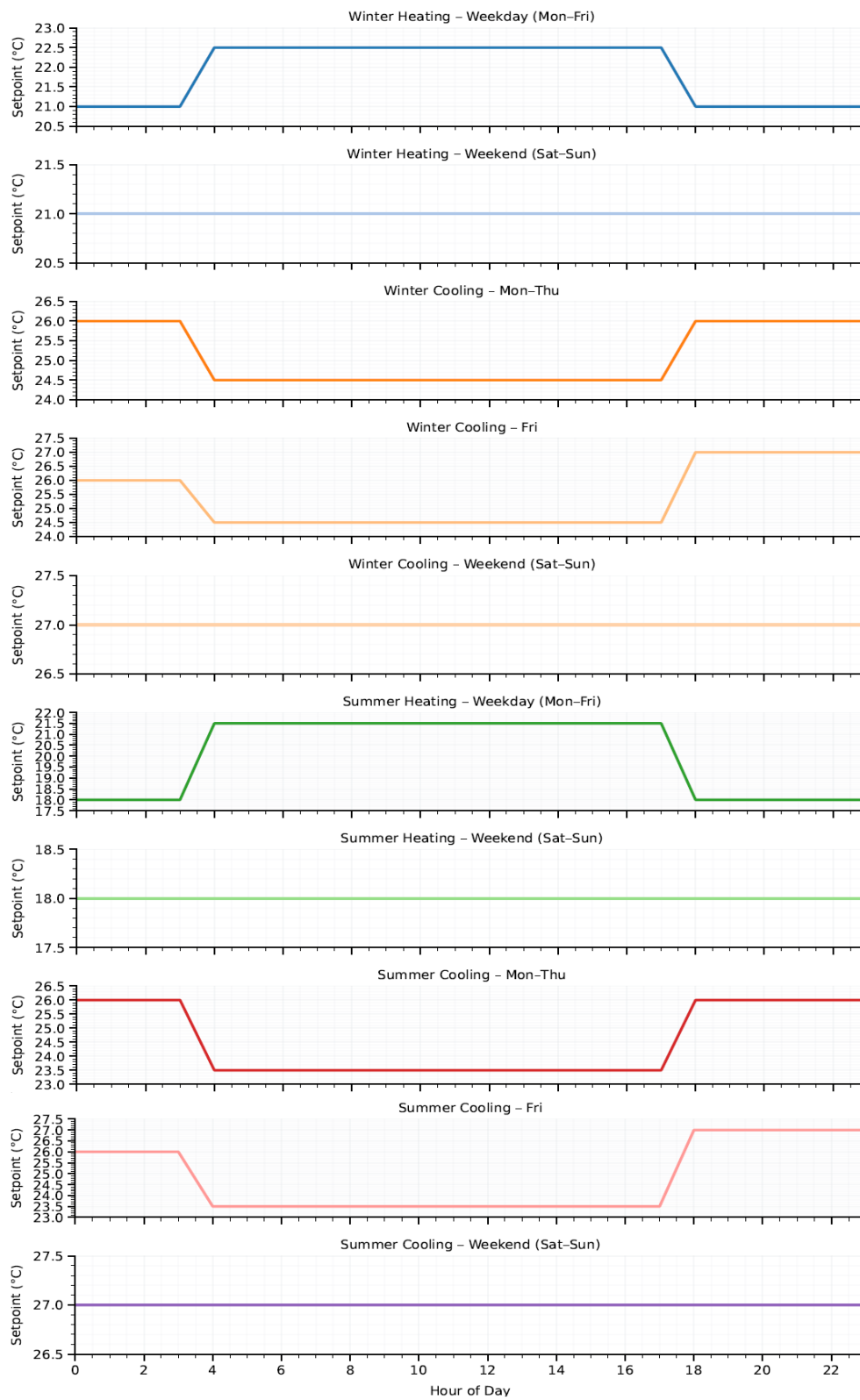


Figure 5-3- HVAC setpoint schedule for parametrical numerical model based on high-performance case study building

Details for mechanical system components are as follows:

- **Primary Equipment:** VRF outdoor heat pump units with capacity control
- **Zone Distribution:** Individual fan-coil terminal units per thermal zone
- **Ventilation:** Dedicated ERV with plate heat exchangers (76-81% effectiveness)
- **Control Strategy:** Dual setpoint temperature schedules with economizer operation
- **Heating COP:** 3.2 at design conditions
- **Cooling COP:** 4.1 at design conditions
- **Supply Air Temperature:** 18-27°C range
- **Ventilation Rate:** 7 L/s/person minimum outdoor air

Schematic of VRF system and ventilation system for detailed model is presented in Figure 5-1 and Figure 5-4

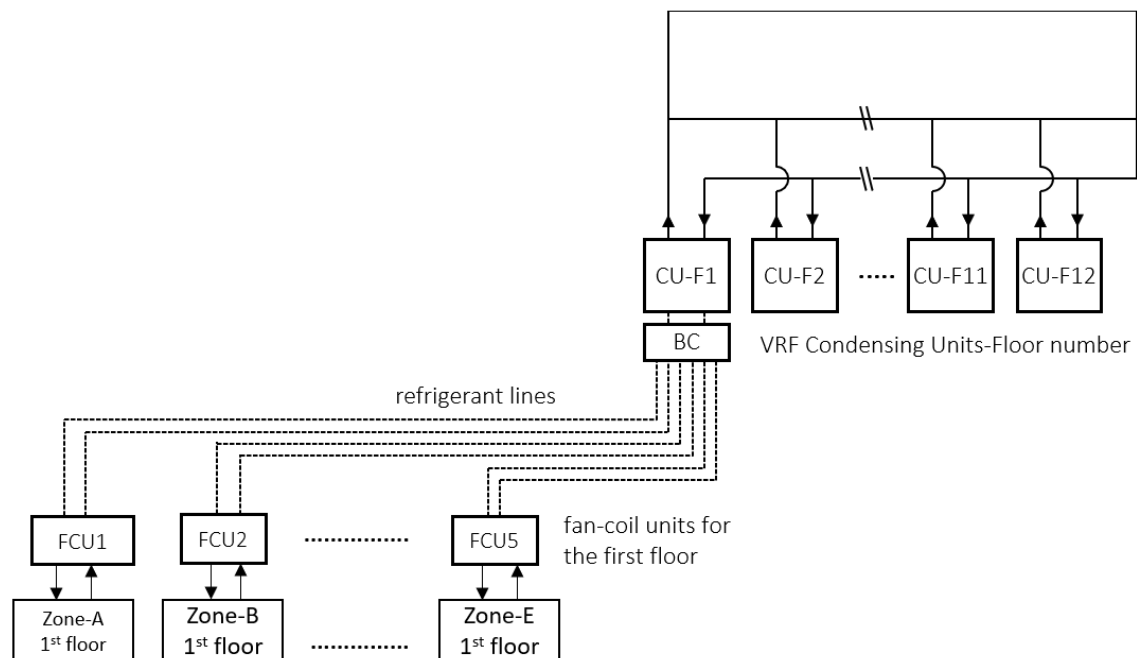


Figure 5-4- Schematic of VRF system for zones (separated terminal for each zone)

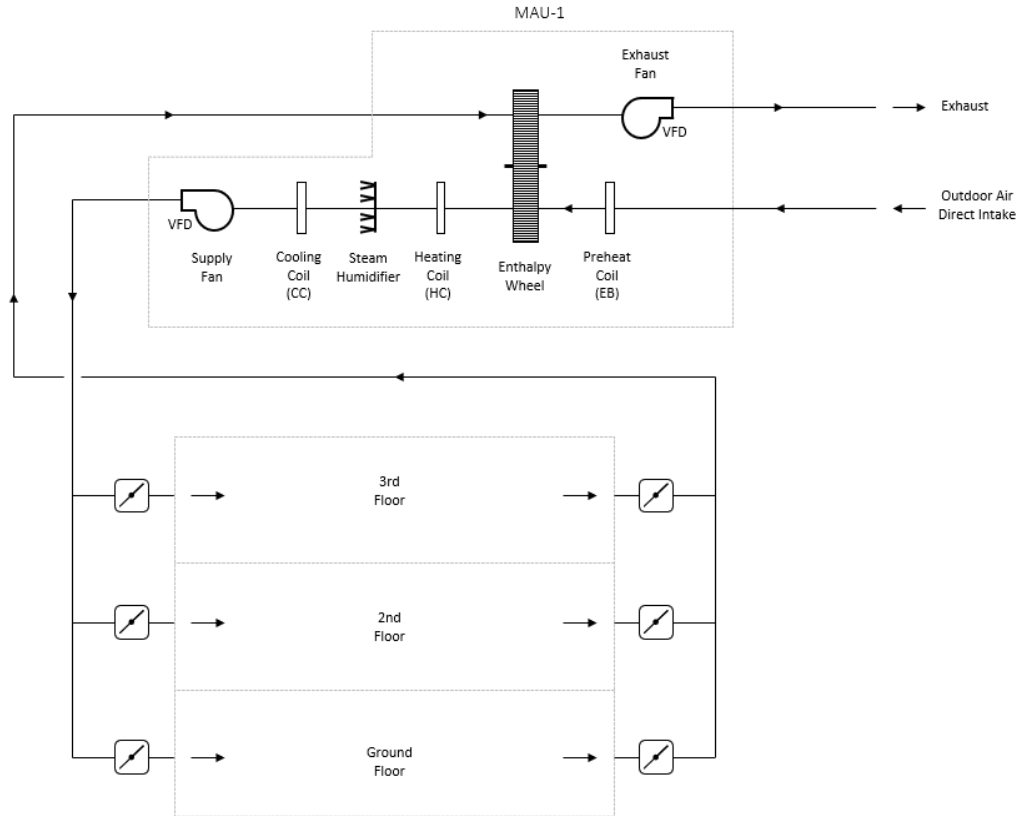


Figure 5-5- Schematic of Energy Recovery Ventilation (ERV) system

5.1.5 Parametric Model Inputs

Given the large number of parameters involved in building energy analysis, a focused subset was selected for systematic variation in the parametric framework. This selection is guided by three considerations: physical significance, practical relevance at the design stage, and compatibility with a continuous parametric sampling strategy. The twelve continuous design parameters selected for systematic variation span building geometry, envelope thermal performance, glazing properties, internal load intensities, and infiltration. Together they capture the principal design decisions available to engineers and architects at the early design stage and provide sufficient variability in the synthetic dataset to support reliable surrogate model training across the defined parameter space.

The 12 main input parameters include:

1. Floor area (A)
2. Rectangular building aspect ratio (L/W)
3. Window-to-wall ratio (WWR)

4. Wall effective thermal resistance
5. Roof effective thermal resistance
6. Infiltration rate
7. Window solar heat gain coefficient (SHGC)
8. Window effective thermal transmittance
9. Lighting power density (LPD)
10. Equipment power density (EPD)
11. Building orientation
12. Number of middle floors (in addition to ground and top floor)

Two categories of parametric variables, simple and geometrical, were considered. This classification was necessary due to the process required to define a parametric variable within the EnergyPlus code for modifying the building model in order to generate the synthetic dataset.

5.1.6 Simple variables

Simple parameters (such as window SHGC, window thermal transmittance, roof and wall thermal resistance, infiltration rate, building orientation, number of middle floors, and internal load intensities) are defined by single values in the EnergyPlus code. Therefore, these parameters can be modified directly within the parametric model by changing their values.

It should be noted that the number of middle floors is handled using the “ZoneMultiplier” class in EnergyPlus, which allows it to be treated as a simple parameter, even though it may appear to be a geometric parameter.

5.1.7 Geometrical parameters

For geometry-related parameters (Area, Aspect Ratio, WWR) a function was implemented to modify vertex coordinates in the EnergyPlus geometry definitions. Changes in area and aspect ratio were applied by updating vertices in the “BuildingSurface:Detailed” objects, whereas WWR modifications were applied to the “FenestrationSurface:Detailed” objects.

5.1.8 Parameter Range

The range of the twelve input parameters was primarily defined to reflect performance levels consistent with NZEB objectives. However, the ranges were intentionally extended beyond these values to also

include characteristics representative of more conventional building practices. The selected ranges are informed by a combination of building codes, literature, and characteristics observed in the case study building presented in Chapter 4.

By expanding the parameter bounds in this way, the study captures a broader design space and ensures that the analysis includes high-performance configurations near the parameter boundaries. While some combinations at the extreme ends of the parameter space may be less common in typical practice, they remain physically plausible. This approach allows the model to explore the full spectrum of possible outcomes and to identify the boundaries of system performance across the 12-dimensional input parameter space. This is particularly important for surrogate model development, where sufficient variability in the training dataset is required to accurately capture nonlinear relationships between input parameters and building energy performance.

- **Aspect ratio and building area:** Only rectangular building geometries are considered, using aspect ratio, r , where $r = L/W$ and the floor area is equal to $A = L \times W$, with dimensions expressed as $L = \sqrt{A \cdot r}$ and $W = \sqrt{A/r}$. To avoid unrealistic geometries, the aspect ratio range is [0.3–1]. As illustrated in Figure 5-6, Considering perimeter zones of 5 m deep on all sides and a minimum core zone of 1 m, the minimum feasible building area is 404 m², and the upper bound is set to 4,000 m², encompassing the case study building of 3,200 m².
- **Building orientation:** Since the building shape is varied by adjusting the aspect ratio, a directional angle range of 0° to 180° from true north is considered in order to capture all possible building orientation scenarios.
- **Number of floors:** According to the Prototype Building Models developed by the U.S. Department of Energy, medium office buildings typically consist of three floors, while large office buildings have twelve floors. Therefore, in the data generation process, the number of floors is varied within the range of 3 to 12.
- **Exterior Wall Thermal Resistance:** Based on NECB 2020, the prescriptive minimum thermal resistance requirements for above grade opaque wall assemblies for Climate Zones 6A is RSI-4.2 [10]. In this study, a slightly lower value of RSI-3.5 was set as the lower limit to include the code minimum value within the LHS range. The upper limit, RSI-8.8, was selected based on the studies on High-Performance Wall Assemblies [140], [141]. Therefore, the wall thermal resistance range considered in this study is [RSI-3.52 to RSI-8.81].
- **Roof Thermal Resistance:** According to NECB 2020 [10], and ASHRAE 90.1-2022 [9] the minimum roof thermal resistance for Climate Zone 6A is RSI-5.6. This study uses RSI-5.3 to include the code minimum value and sets RSI-10.6, based on the studies on high-performance

assemblies [140], [141]. Therefore, the wall thermal resistance range considered in this study is [RSI-5.28 to RSI-10.58].

- **Window Properties:** A window thermal transmittance range between 0.57-1.95 W/m²·K is selected to represent double and triple glazing configurations, aligning with current high-performance standards. The Solar Heat Gain Coefficient (SHGC) varies from 0.25–0.70, covering tinted and reflective glass but also exterior shading in the form of an equivalent SHGC.
- **WWR:** A study by Goia (2016) found that the optimal WWR for office buildings in various European climates ranged between 30% and 45% [142]. In Canada, the NECB sets a maximum WWR of 40%, and Ontario's SB-10 standard aligns with this limit, requiring performance-based compliance for higher ratios [143][144]. Therefore, the WWR range selected for this study is [15%–55%] consider lower range for lower energy consumption as well.

All windows are set to 0.8 m above the related floor, WWR is adjusted by changing the height of the windows which is applied by editing the Z axis coordinate of the window sill. It should be noted that WWR remains the same for all the building façade orientations.

- **Lighting power density (LPD):** According to ASHRAE 90.1 using the Building Area method, the recommended LPD for office buildings is 6.7 W/m² [9]. In this study, a range of [2.0–8.0] W/m² is selected. The lower bound (2.0 W/m²) is intentionally set to reflect rapid advances in LED efficacy, luminaire performance, and lighting controls, and to capture high-performance and future-ready lighting designs that can achieve substantially lower LPD than today's prescriptive values. The upper bound (8.0 W/m²) provides coverage for more conventional or feature-intensive lighting layouts while remaining within a plausible envelope for office applications. Overall, the selected range spans both current practice and anticipated near-future lighting solutions, while staying consistent with typical regulatory and design constraints.

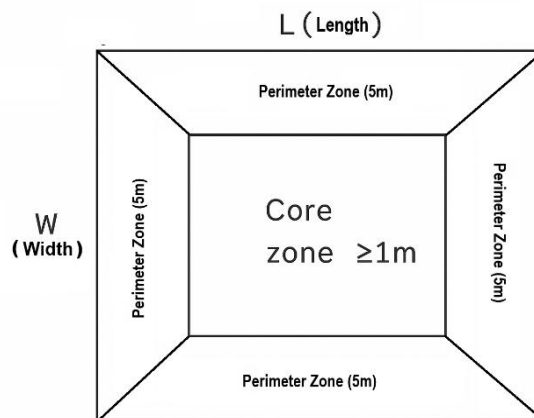


Figure 5-6- Floor zoning scheme defining the minimum building dimensions and resulting minimum floor area based on the aspect ratio.

- **Equipment power density (EPD):** EPD represents plug load intensity, accounting for computers, appliances, and office equipment. Typical values for office spaces range from 5–15 W/m², depending on equipment density and operational hours. To cover both conventional and high-performance offices, this study adopts a range of 2–20 W/m². Lower bounds represent highly efficient offices equipment, while upper bounds capture equipment-intensive zones such as servers.
- **Infiltration:** Based on ranges reported in the literature, building airtightness values typically vary between 0.1 and 1.0 m³/h·m² depending on construction quality and envelope detailing[145], [146]. This range was adopted in this study to represent plausible infiltration conditions in the parametric simulations. For implementation in EnergyPlus, where SI units are required, this corresponds to a range of 0.000028 to 0.00028 m³/s·m².

Table 5-1 summarizes the input parameters details which are explained in the previous sections:

Table 5-1- Input parameters, range, and type of parameters.

Parameter (unit)	Min	Max	Type
Building Footprint Area (m ²)	404	4000	geometry
Floor Plate Aspect Ratio; L/W (-)	0.3	1	geometry
Building Orientation Relative to True North (°)	0	180	simple
Number of middle floors; N_of_Fl (floors, integer)	1	10	simple
Wall Thermal Resistance; Wall-R (m ² ·K/W)	3.52	8.81	simple
Roof Thermal Resistance; Roof-R (m ² ·K/W)	5.28	10.58	simple
Window Thermal Conductance; U-factor (W/m ² ·K)	0.57	1.95	simple
Window Solar Heat Gain Coefficient; SHGC (-)	0.25	0.7	simple
Window-to-Wall-Ratio; WWR (-)	0.15	0.55	geometry
Infiltration per Exterior Wall Area (m ³ /s·m ²)	2.8×10 ⁻⁵	3.0×10 ⁻⁴	simple
Lighting Power Density; LPD (W/m ²)	2	8	simple
Equipment Power Density; EPD (W/m ²)	3	20	simple

5.1.9 Weather Input

The weather conditions were represented using the TMY3 weather file for Waterloo Region, Ontario, Canada (“CAN_ON_Region.of.Waterloo.Intl.AP.713680_TMYx.2007-2021.epw”; ASHRAE climate zone 6A), which was used as a fixed input to the numerical model. The file contains hourly meteorological data, including outdoor air temperature, relative humidity, solar radiation, wind speed, and other relevant parameters necessary for building energy modelling.

5.1.10 Simulation Time Steps

All simulations were performed using 10-minute intervals (6 timesteps per hour) to accurately capture building thermal and system dynamics. For surrogate modelling purposes, results were aggregated and reported at an hourly resolution.

5.1.11 Model Outputs

After completing all model setup procedures, the EnergyPlus simulations were executed to generate the output variables that constitute the synthetic data used for the development and evaluation of the surrogate model. Only the output variables directly used in this study were extracted and retained, limited to:

- Hourly total electricity consumption (used as the primary high-resolution output for surrogate modelling)
- Annual energy performance indicators, including:
 - Energy Use Intensity (EUI)
 - Thermal Energy Demand Intensity (TEDI)
 - Cooling Energy Demand Intensity (CEDI)

Other simulation outputs (e.g., zone-level temperatures, humidity, or system-level diagnostics) were used solely during model development and verification but were not included in the final dataset for surrogate model training or evaluation.

5.1.12 Simulation Failure Handling

Automated quality-control checks were implemented to ensure dataset reliability. Simulations producing severe or fatal EnergyPlus errors were flagged and excluded. Models exhibiting non-convergent behavior or physically inconsistent results were removed from the dataset. Only successfully completed valid simulations were retained for surrogate modelling.

5.1.13 Python-Based Automation and Validation

All simulations were executed using a Python-based automation framework built on the EnergyPlus Python API. The workflow included programmatic modification of IDF files, execution of simulations, validation of modified inputs, and structured extraction of hourly and annual outputs. This approach ensured repeatability, consistency, and traceability across the large set of parametric simulations.

5.1.14 Computational Environment

Simulations were performed on a single CPU-based workstation. Batch execution and efficient data handling strategies were used to manage computational load.

5.2 Numerical Stability and Indoor Environmental Response

This first step aims to verify that the custom parametric EnergyPlus model executes reliably and produces physically plausible indoor environmental conditions for randomly sampled parameter combinations. The purpose is functional verification (quality assurance of implementation and control logic), rather than calibration to measured data. Simulation diagnostics confirm numerical stability and successful model execution. The EnergyPlus simulation is converged with no severe error and the results shows that the complete time series output (i.e., 8,760 hours per year) is generated. Internally, the simulation used six timesteps per hour, which served as the base temporal resolution, while hourly data were retained for analysis.

Note that a 10-minute timestep (i.e., 6 timesteps/hour in EnergyPlus) is commonly used because it is short enough to accurately resolve thermal storage dynamics, HVAC control response, and latent moisture processes, yet long enough to keep simulations computationally efficient.

HVAC operation is governed by schedule-based heating and cooling setpoints. During the winter schedule, a minimum zone temperature setpoint of 21 °C is maintained after 18:00 on weekdays and throughout weekends as part of setback operation. A maximum zone temperature limit of 27 °C is also defined during winter to allow cooling if indoor temperatures exceed acceptable levels. These values represent the lower and upper bounds of the zone temperature during unoccupied winter periods.

Similarly, during the summer schedule, a minimum zone temperature setpoint of 18 °C is applied after 18:00 on weekdays and throughout weekends to permit heating if temperatures drop excessively. A maximum zone temperature limit of 27 °C is enforced during unoccupied summer periods to reduce cooling energy consumption while maintaining acceptable indoor conditions. Occupied-hour setpoints differ from these setback limits and are presented in Figure 5-3.

Hourly zone mean air temperature and operative temperature were evaluated for all zones to verify correct thermostat control behavior. Across the simulation period, the minimum recorded temperature was

17.51 °C and the maximum was 27.6 °C, under Waterloo weather input data. These extrema remain within approximately ± 1 °C of the scheduled setpoints, which is consistent with expected HVAC control tolerances and indicates that setpoint scheduling and system response are functioning as intended.

Figure 5-7 and Figure 5-9 present hourly setpoint temperatures (red), mean zone air temperature (blue) and mean operative zone temperature (green) for one representative winter week and one representative summer week, respectively. The corresponding outdoor conditions for these weeks are provided in the companion figure. The weekly profiles show clear transitions between occupied and setback periods and remain bounded within the expected temperature range, confirming correct implementation of the time-of-day and day-of-week HVAC schedules as well as stable heating/cooling operation.

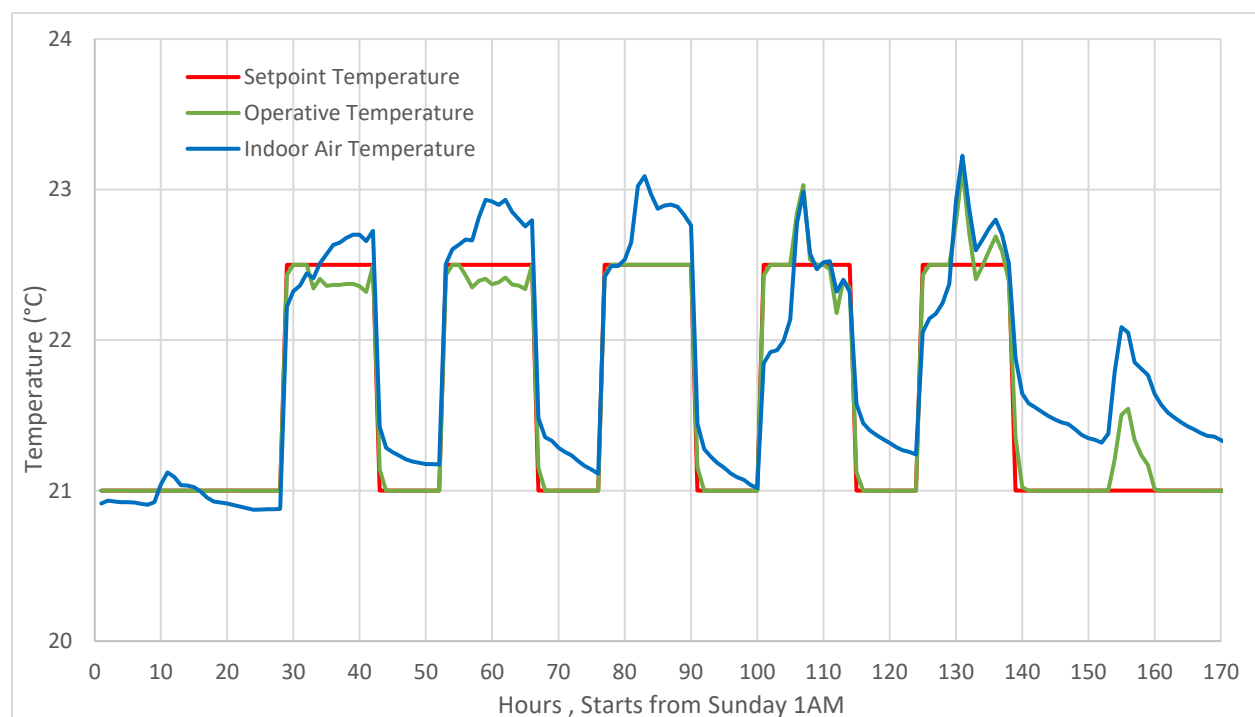


Figure 5-7- One representative week of hourly setpoint and operative temperature for one of the east perimeter zones during wintertime (heating season).

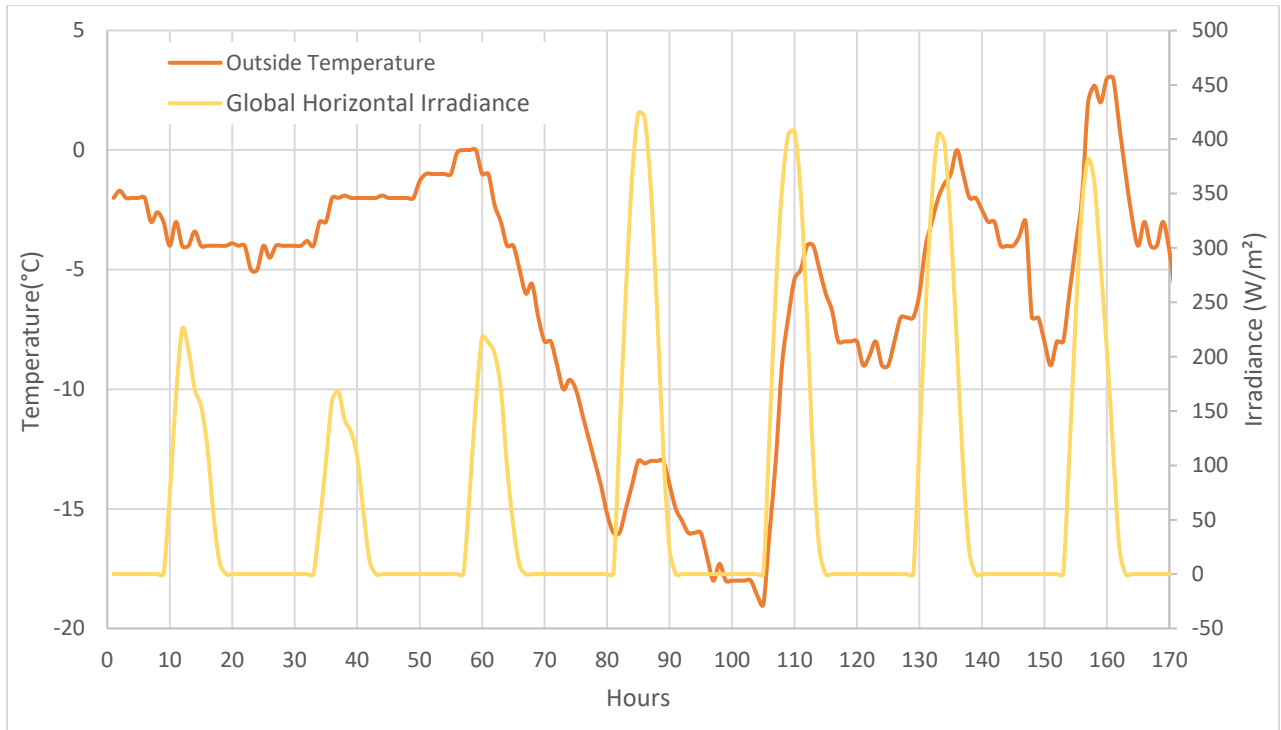


Figure 5-8- Outdoor temperature and global horizontal irradiance for the representative winter week

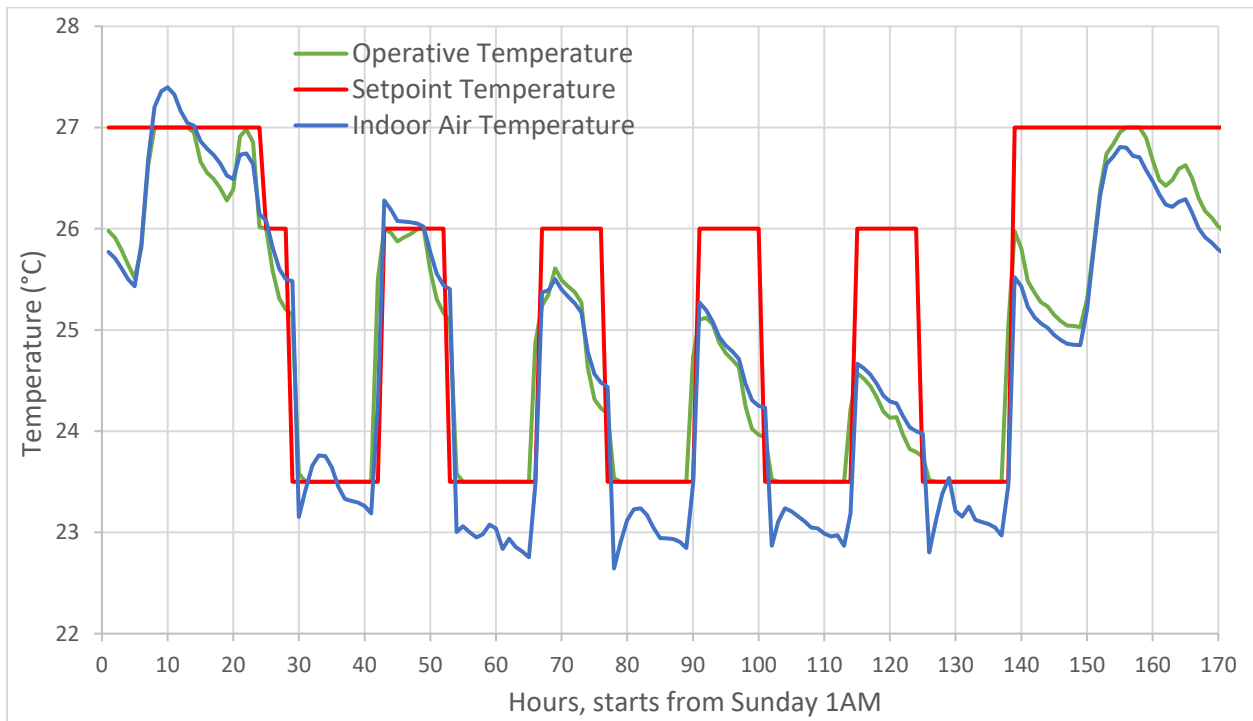


Figure 5-9- One representative week of hourly setpoint and operative temperature for one of the east perimeter zones during summertime (cooling season).

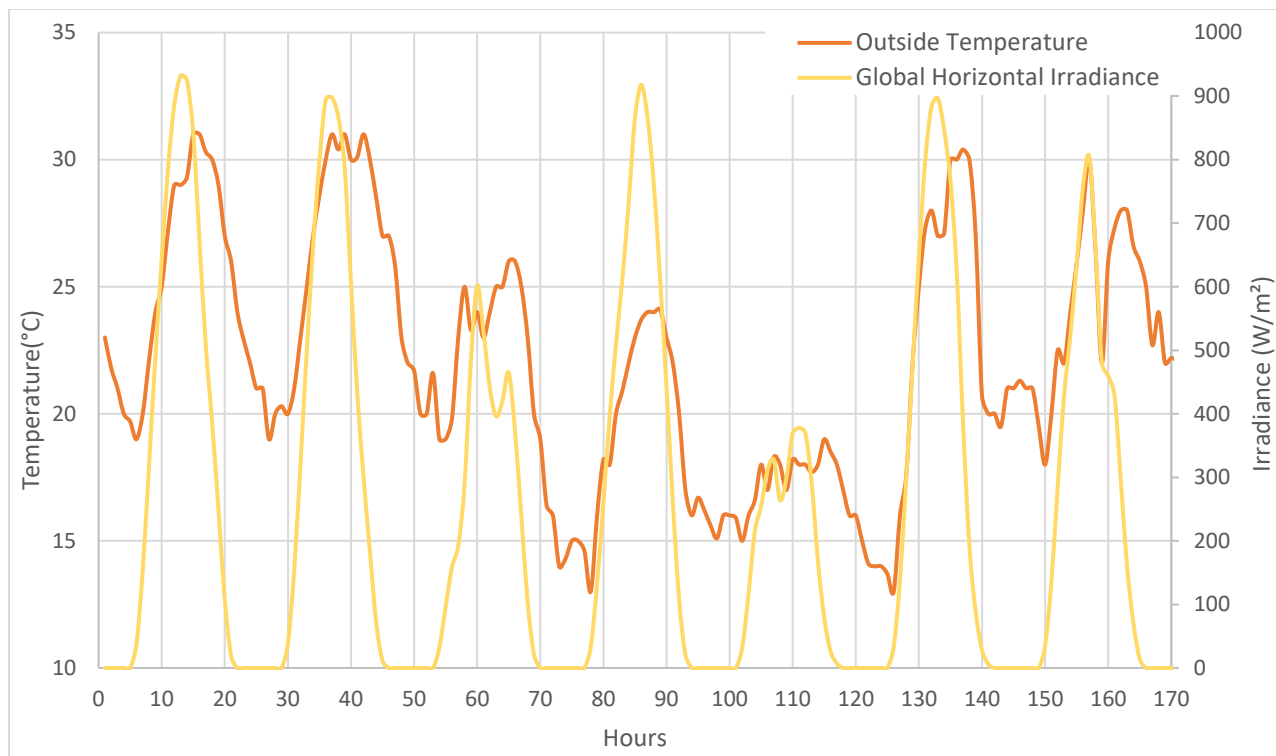


Figure 5-10- Outdoor temperature and global horizontal irradiance for the representative summer week

5.3 Parametric Sensitivity Analysis of the Numerical Energy Model

To evaluate the parametric detailed building energy model, the influence of each input parameter on the key performance indicators, EUI, TEDI, and CEDI, were examined. The objective is not calibration, but behavioral verification; to confirm that the model produces logical and physically interpretable trends across the defined parameter space before it is used for surrogate model training. It also provides an early understanding of how each parameter shapes energy performance prior to surrogate model development.

The sensitivity analyses are conducted using a one-at-a-time (OAT) approach, in which individual parameters are varied systematically while all other parameters are held constant. This allows for the isolation of the effect of each parameter and enables clear interpretation of cause-effect relationships.

Two types of sensitivity analyses are performed. First, a worst-case (extreme scenario) analysis is conducted by varying parameters across their full defined ranges to evaluate the bounds of system performance. Second, a case study–based sensitivity analysis is carried out using baseline values corresponding to the case study building, allowing assessment of parameter influence under realistic operating conditions.

Twelve input parameters were considered, spanning geometry, envelope properties, internal loads, orientation, and air leakage. Their ranges were summarized in Table 5-2.

Table 5-2- Input parameters base scenario value and their range

Parameter name or explanation (Unit)	Base value	Range
Building Footprint Area (m ²)	404	404 to 4000
Floor Plate Aspect Ratio; L/W (-)	0.3	0.3 to 1
Window Thermal Conductance; U-factor (W/m ² ·K)	1.95	1.95 to 0.57
Window Solar Heat Gain Coefficient; SHGC (-)	0.7	0.7 to 0.25
Building Orientation Relative to True North (°)	0°	0°-180°
Wall Thermal Resistance; Wall-R (m ² ·K/W)	3.52	3.52 to 8.81
Roof Thermal Resistance; Roof-R (m ² ·K/W)	5.28	5.28 to 10.67
Infiltration per Exterior Wall Area (m ³ /s·m ²)	3×10^{-4}	3×10^{-4} to 2.8×10^{-5}
Window-to-Wall-Ratio; WWR (-)	0.55	0.55 to 0.15
Lighting Power Density; LPD (W/m ²)	8	8 to 2
Equipment Power Density; EPD (W/m ²)	20	20 to 3
N_of_Fl (Number of middle floors)	1	1-10

5.3.1 Sensitivity Analysis for the Worst-Case Scenario

A *worst-case* scenario (refer to as baseline, hereinafter) was first established by assigning each input parameter the min or max value within its range, resulting in producing the highest overall energy demand. This scenario serves as a consistent baseline from which individual parameter effects can be evaluated. Each parameter range was then discretized into five values, including the baseline value. For the integer variable representing number of middle floors (*N_of_Fl*), the closest integer value was chosen at each subdivision; therefore, its normalized x-axis spacing appears non-uniform in the plots.

Using this baseline, each parameter was varied independently across its allowable range while all others were held constant. The resulting impacts were evaluated for three annual performance indicators:

- Energy Use Intensity (EUI),
- Thermal Energy Demand Intensity (TEDI), and
- Cooling Energy Demand Intensity (CEDI).

The corresponding response curves are shown in Figure 5-11, Figure 5-12, and Figure 5-13, respectively.

5.3.2 Hypotheses

To formally assess the physical consistency of the detailed numerical model, the following hypotheses were defined based on established building-physics principles:

- H1: For a rectangular building with symmetric geometry and constructions, a 90° rotation of the building around the vertical axis should produce similar EUI, TEDI, and CEDI values.
- H2: For buildings modeled with fixed perimeter-zone depth and a central core zone, increasing the footprint area, or increasing aspect ratio in a way that increases core-zone floor area relative to perimeter-zone area, will reduce EUI.
- H3: Increasing the number of floors reduces EUI due to a decreasing envelope-to-floor-area ratio, while the magnitude of this reduction diminishes as the number of floors increases.
- H4: In cooling-dominated buildings, improving insulation reduces TEDI but may increase CEDI.
- H5: Reducing SHGC increases TEDI and reduces CEDI.
- H6: Increasing air leakage increases EUI and TEDI and reduces CEDI.
- H7: Increasing internal gains, including EPD and LPD, increases EUI and CEDI and reduces TEDI.
- H8: Insulation has diminishing influence on total energy consumption under the modeled internal load conditions, even though it affects TEDI and CEDI individually.

Agreement between simulation outputs and these expected trends is taken as evidence that the model behaves consistently with physical principles.

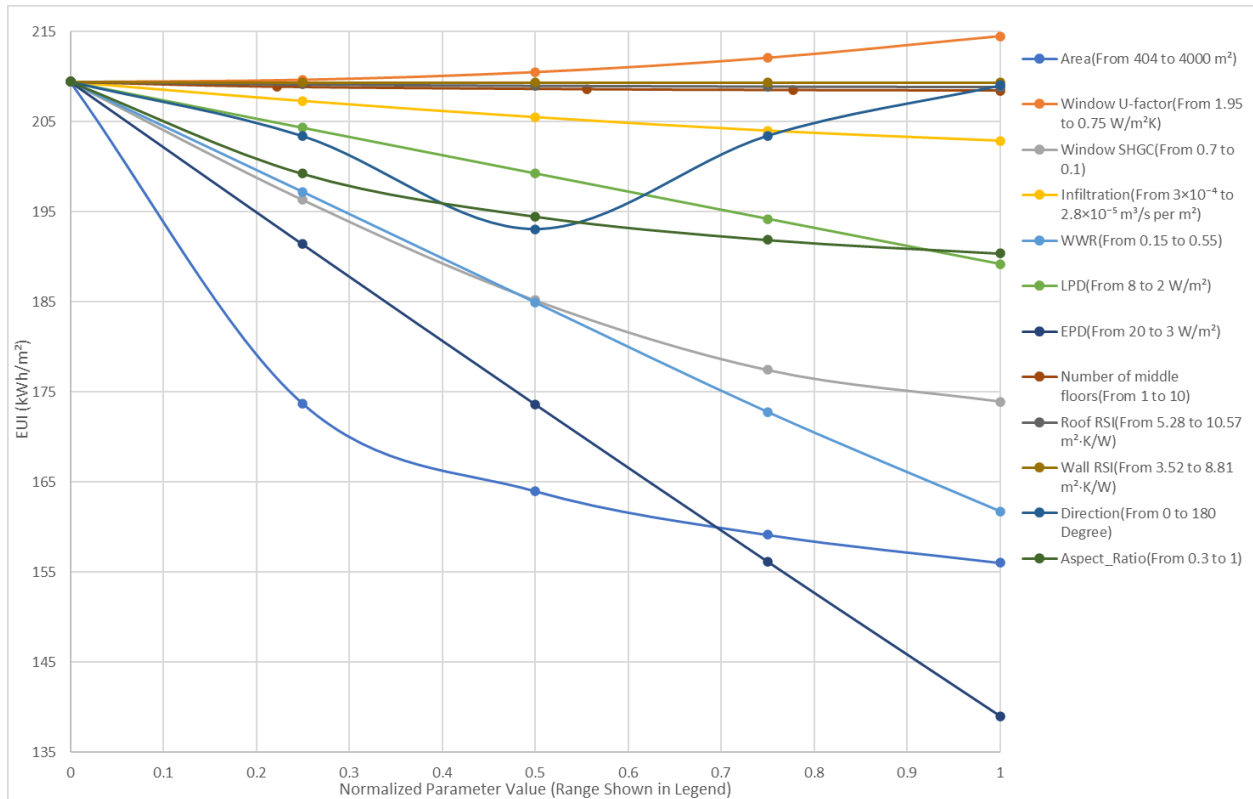


Figure 5-11- EUI for different scenarios, based on the total energy in the worst case.

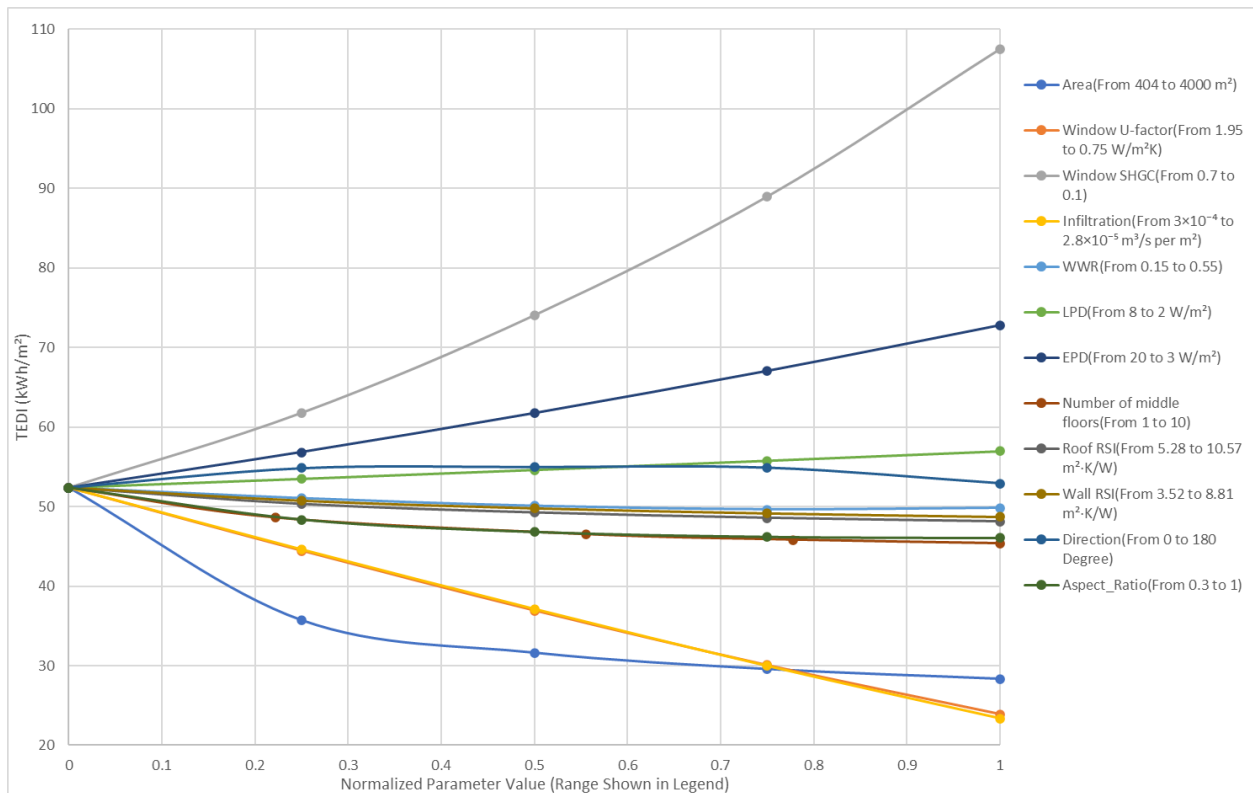


Figure 5-12-TEDI for different scenarios, based on the total energy in the worst case

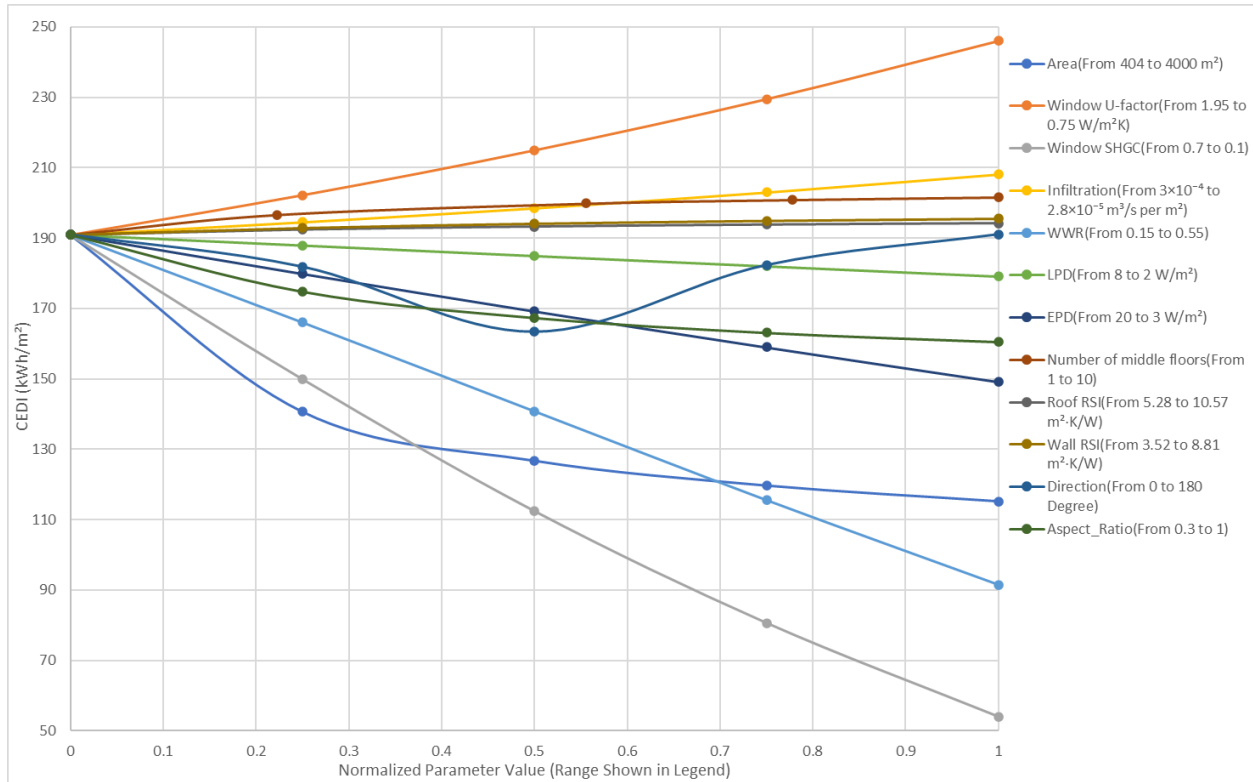


Figure 5-13-CEDI for different scenarios, based on the total energy in the worst case

5.3.3 Hypothesis Verification

The verification of the predefined hypotheses is based on the sensitivity results presented in Figure 5-11, Figure 5-12, and Figure 5-13, which illustrate variations in EUI, TEDI, and CEDI respectively across the parameter space under worst-case total energy conditions.

5.3.3.1 H1: Orientation Symmetry

As shown in Figure 5-11 to Figure 5-13 the oscillatory variation observed for building direction confirms that solar gains vary cyclically with rotation. A 90° rotation produces comparable EUI, TEDI, and CEDI values, with only minor deviations due to seasonal solar distribution differences. Therefore, H1 is satisfied, indicating correct implementation of solar geometry and façade exposure effects.

5.3.3.2 H2: Footprint Area and Aspect Ratio

Simulation results show that increasing footprint area generally reduces EUI, particularly when the core-zone area increases relative to perimeter zones. This behavior is consistent with reduced façade exposure per unit floor area. Aspect ratio variations produce moderate changes, reflecting redistributed solar exposure but not contradicting the expected trend.

5.3.3.3 H3: Number of Floors

Increasing the number of floors reduces EUI, primarily due to the decreasing envelope-to-floor-area ratio. However, the marginal reduction becomes smaller at higher floor counts, demonstrating diminishing returns. This confirms H3 and indicates correct representation of vertical compaction effects.

5.3.3.4 H4: Insulation Effects Window thermal transmittance and Opaque envelope thermal resistance)

Reducing window U-factor decreases TEDI while increasing CEDI. Because the base scenario uses high EPD and LPD, the building behaves as cooling-dominant, and the increase in cooling demand outweighs the heating reduction, leading to a net increase in EUI. This confirms H4 and is physically consistent with internal-gain-dominated buildings.

For wall and roof thermal resistance, changes produce minimal impact on EUI and only modest effects on TEDI and CEDI. This behavior reflects diminishing returns of insulation improvements in cooling-dominated buildings with high internal loads. The hypothesis is therefore conditionally supported, with the limited magnitude explained by internal load dominance and already efficient baseline envelope performance.

5.3.3.5 H5: Solar Heat Gain Coefficient

Reducing SHGC leads to significant reductions in CEDI and a slight increase in TEDI. The net effect generally reduces EUI, confirming the dominant role of solar gains in cooling-dominated applications. Therefore, H5 is satisfied.

5.3.3.6 H6: Infiltration

Increasing infiltration raises TEDI due to higher uncontrolled outdoor air exchange and moderately increases EUI. Its effect on CEDI is smaller and may vary depending on seasonal conditions. Overall, the simulation trends align with expected physical behavior, and H6 is supported.

5.3.3.7 H7: Internal Gains (EPD and LPD)

Increasing EPD and LPD produces substantial increases in EUI and CEDI while reducing TEDI, consistent with increased internal heat gains offsetting heating demand and intensifying cooling loads. The magnitude of the effect confirms that internal gains are dominant drivers in this building type. Therefore, H7 is clearly satisfied.

5.3.3.8 H8: Wall and Roof Thermal Resistance

A sensitivity analysis was performed to examine the effect of envelope thermal resistance on building energy performance. Wall and roof thermal resistances were varied from RSI-0.88 to RSI-7.0 while all other parameters were held at their baseline values (Figure 5-14 to Figure 5-16).

The results show that the modeled building is predominantly cooling-dominated under the assumed high internal loads (worst-case LPD and EPD), despite being located in a cold climate. As envelope thermal resistance increases, TEDI decreases because heat loss through the envelope is reduced (Figure 5-15). In contrast, CEDI increases because greater insulation limits heat dissipation, thereby increasing cooling demand during occupied periods (Figure 5-16).

These opposing effects largely offset one another, resulting in only modest changes in total EUI across the range of thermal resistance considered. The sensitivity of TEDI, CEDI, and EUI is greatest at lower insulation levels ($RSI < 3.5$). Beyond approximately $RSI-3.5$, further increases in insulation produce only small changes in the energy indicators, indicating diminishing sensitivity for this building configuration.

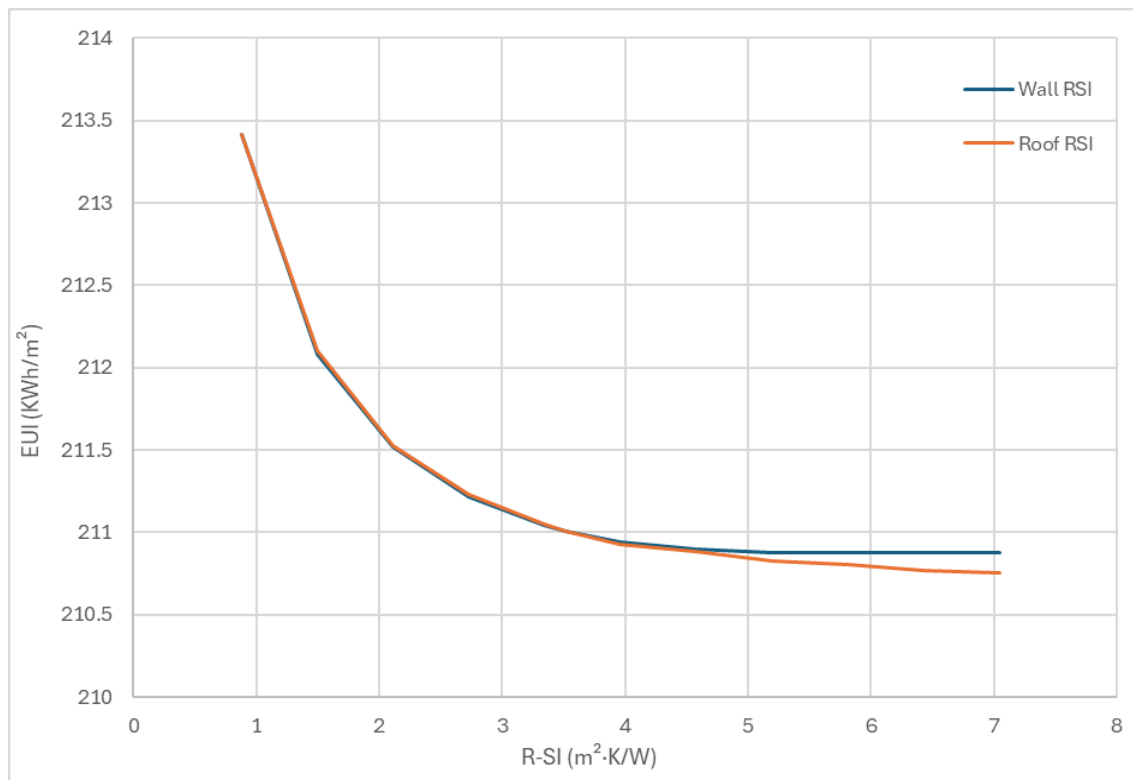


Figure 5-14- Sensitivity of EUI to variations in wall and roof thermal resistance, with all other variables fixed at baseline values under worse-case conditions,

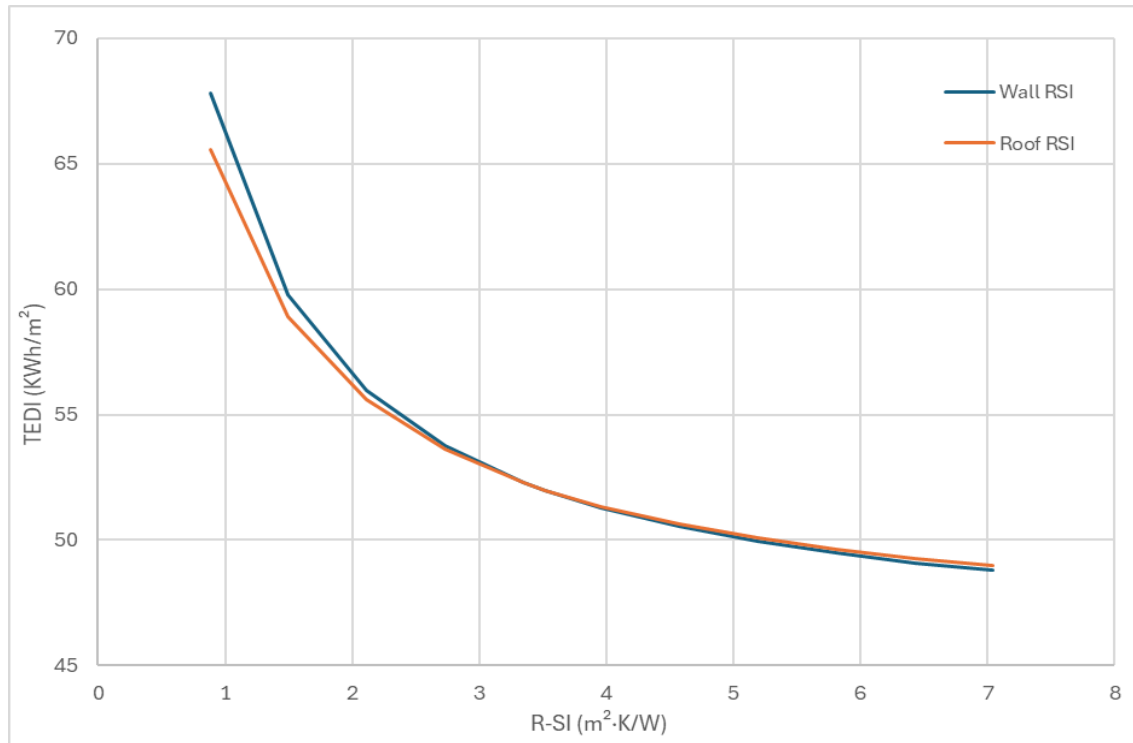


Figure 5-15- Sensitivity of TEDI to variations in wall and roof thermal resistance, with all other variables fixed at baseline values under worse-case conditions.

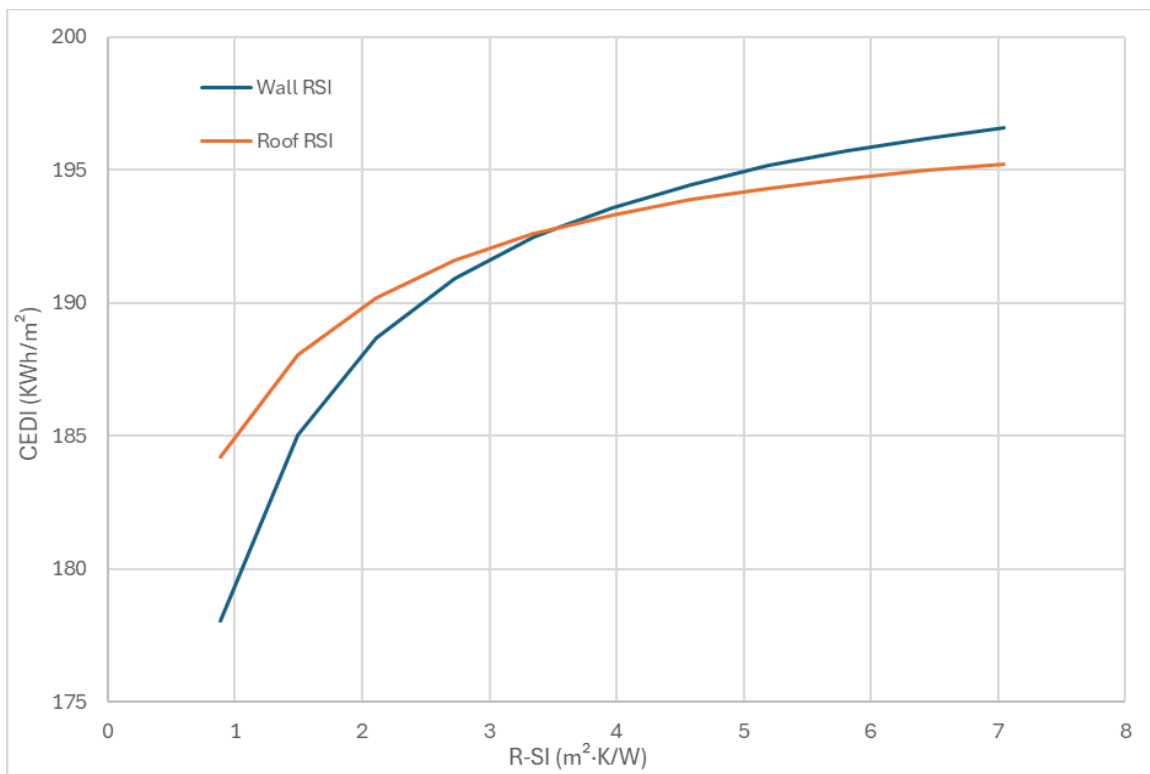


Figure 5-16- Sensitivity of CEDI to variations in wall and roof thermal resistance, with all other variables fixed at baseline values under worse-case conditions.

5.3.4 Sensitivity Analysis for the Office Case Study

To investigate parameter sensitivity for a configuration representative of the case study office building, a specific set of input values was defined. Specifically, parameters describing the physical characteristics of the building, including footprint area, aspect ratio, number of floors, and envelope thermal resistances, were assigned according to the documented properties of the case study building. For parameters that could not be directly determined from the available data, such as infiltration rate and equipment power density, values were selected within physically reasonable ranges and adjusted through a manual model calibration to reproduce the measured EUI of the building. This combination of measured physical attributes and calibrated operational parameters provides a representative input configuration that approximates the case study building within the defined parametric model space and serves as the basis for the sensitivity analysis presented in this section.

This sensitivity analysis focuses exclusively on design parameters that may change during the operational life of the building, either through operation and maintenance practices or as a result of performance degradation over time. Parameters that are inherently fixed once the building is constructed, such as aspect ratio, WWR and orientation were not considered in this analysis. Instead, the assessment concentrates on variables that may realistically vary throughout the building lifecycle, including internal loads, envelope performance characteristics, and infiltration. Evaluating the sensitivity of energy performance to these parameters provides insight into which operational or aging-related changes could most significantly affect the building's ability to maintain net-zero energy performance.

The sensitivity analysis examines how deviations in key design parameters affect the EUI relative to the baseline high-performance office building configuration (Figure 5-17). The baseline case yields an EUI of approximately 67 kWh/m²·yr, which is well below the CGBC net-zero threshold of 100 kWh/m²·yr. The extent to which individual parameters cause the EUI to approach or exceed this threshold provides an indication of the robustness of the building's net-zero performance.

Among the parameters examined, EPD is the dominant driver of EUI. Increasing EPD from the baseline value of 5.5 W/m² to 20 W/m² raises the EUI to approximately 128 kWh/m²·yr, exceeding the CGBC threshold. This result highlights the strong influence of plug loads and IT-related equipment in office buildings. In practice, EPD may increase over time as a result of additional workstation equipment, expanded server or data infrastructure, higher tenant equipment density, and the increased use of personal electronic devices. Because plug loads also generate internal heat gains, they contribute not only to higher electricity consumption but also moderate reduction of heating demand (Figure 5-18) and significant increase of cooling demand (Figure 5-19), thereby amplifying their overall impact on building energy use.

Lighting Power Density (LPD) also has a strong effect on EUI. Increasing lighting density from approximately 3.72 W/m² to 8 W/m² results in a noticeable increase in EUI (Figure 5-17). As with equipment loads, lighting affects building performance through both direct electricity use and added internal heat gains, which in turn increase cooling demand (Figure 5-19). Although lighting technologies generally become more efficient over time with improvements in LED systems and smart controls, LPD may still increase in practice because of changes in space use, higher lighting requirements, longer operating schedules, or adjustments to control settings that reduce the effectiveness of the original design.

Several other design parameters have a moderate effect on EUI but are unlikely to push the building beyond the CGBC threshold within the ranges considered. Window thermal performance has a measurable influence on EUI, although even under poorer glazing conditions the building remains below the EUI limit. Similarly, higher infiltration rates increase EUI by adding to both heating and cooling loads; however, their overall effect is smaller than that of internal loads.

By contrast, wall and roof thermal resistance exhibit relatively low sensitivity with respect to total EUI. Although changes in insulation levels have a clear influence on heating demand, their effect on total energy use is moderated by opposing changes in cooling demand. As a result, variations in wall and roof thermal resistances within the tested range produce only small changes in EUI and do not significantly affect the net-zero energy outcome.

The sensitivity analysis indicates that the net-zero performance of the building is governed more strongly by internal loads than by envelope characteristics. Plug loads and lighting account for the largest variations in EUI, whereas envelope-related parameters have comparatively limited influence. This suggests that maintaining net-zero performance over the building lifecycle depends primarily on controlling operational loads and occupant-related energy use. Moderate degradation in insulation, glazing performance, or airtightness is unlikely to cause the building to exceed CGBC thresholds, whereas uncontrolled increases in equipment and lighting loads present a substantially greater risk to maintaining net-zero performance.

The sensitivity results for the case study configuration are consistent with the physical behavior confirmed in the worst-case analysis, reinforcing confidence in the parametric model's ability to produce physically meaningful outputs across a wide range of design configurations. Together, the numerical stability verification, hypothesis-based worst-case analysis, and case study sensitivity analysis establish that the model responds correctly to all twelve input parameters and captures the expected directional relationships between design decisions and energy performance indicators. This provides the necessary assurance that the synthetic dataset generated from this model (i.e., 3,000 LHS-sampled simulations

producing hourly electricity profiles and annual performance indicators) reflects valid building energy behavior rather than modelling artifacts. The development of surrogate models trained on this dataset is presented in the following chapter, Chapter 6.

The agreement between simulated trends, case study performance indicators, and expected physical behavior further supports the suitability of the parametric model for generating reliable synthetic data for surrogate model development.

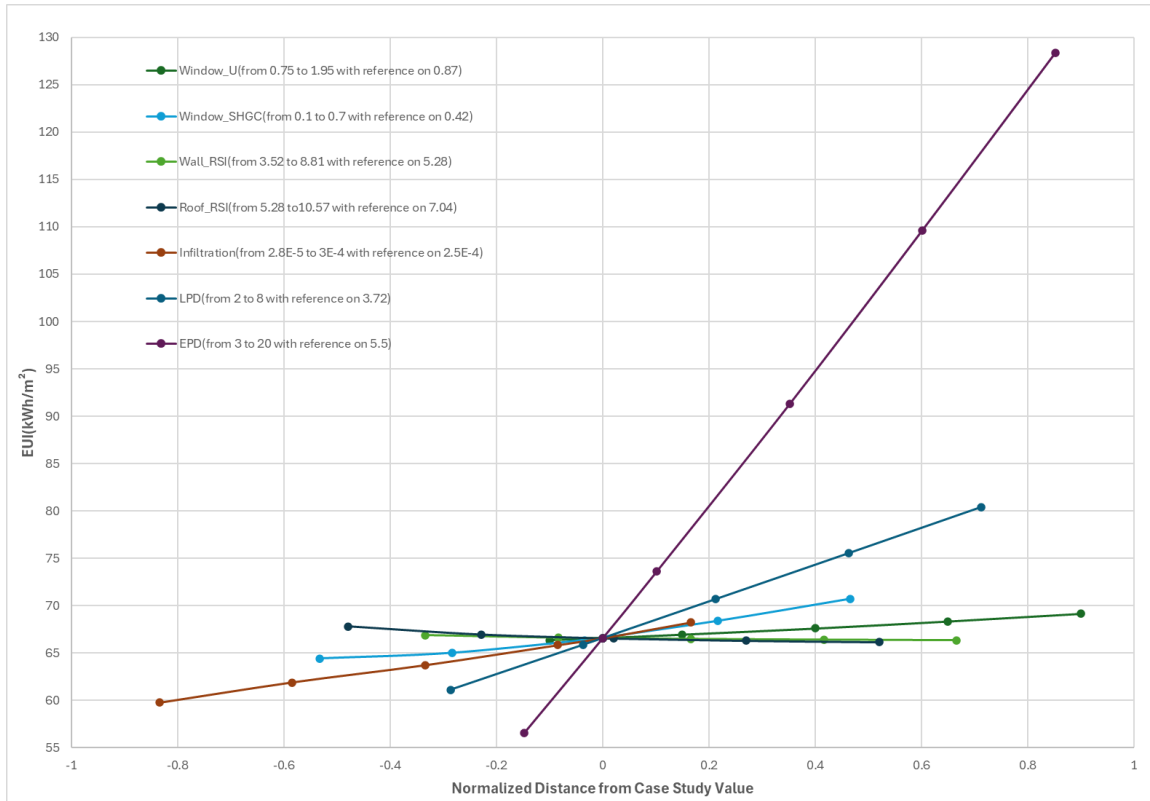


Figure 5-17- Design parameter sensitivity analysis on EUI

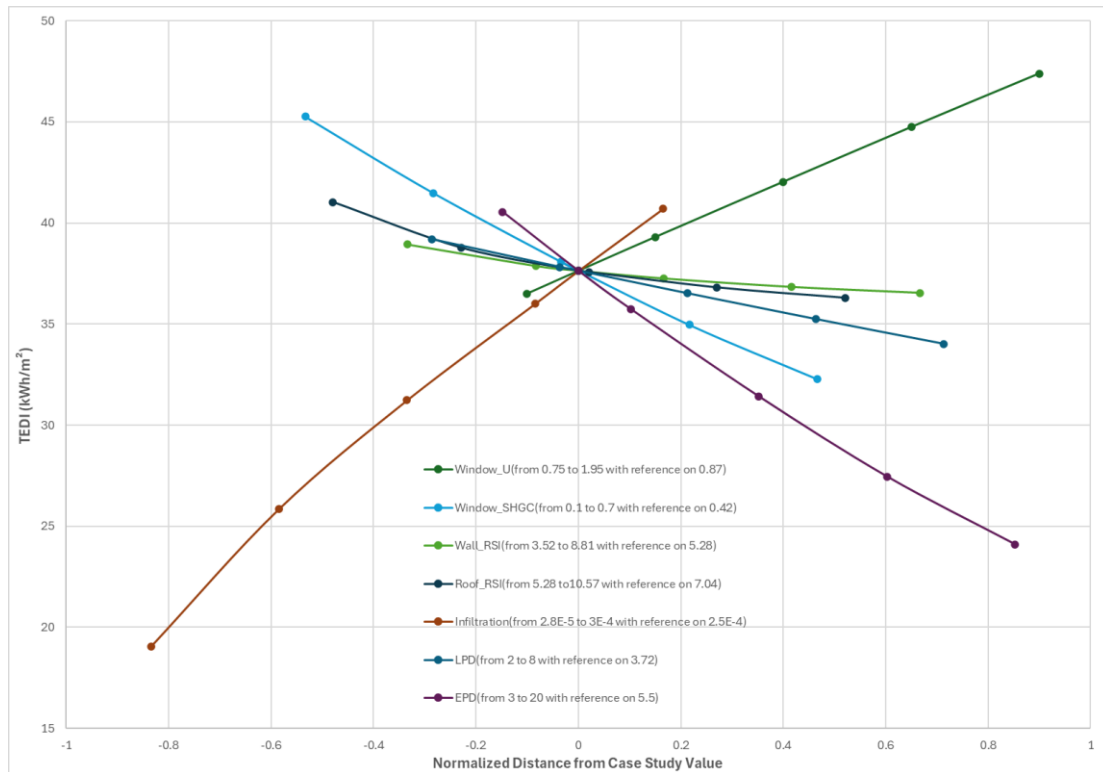


Figure 5-18- Design parameter sensitivity analysis on TEDI

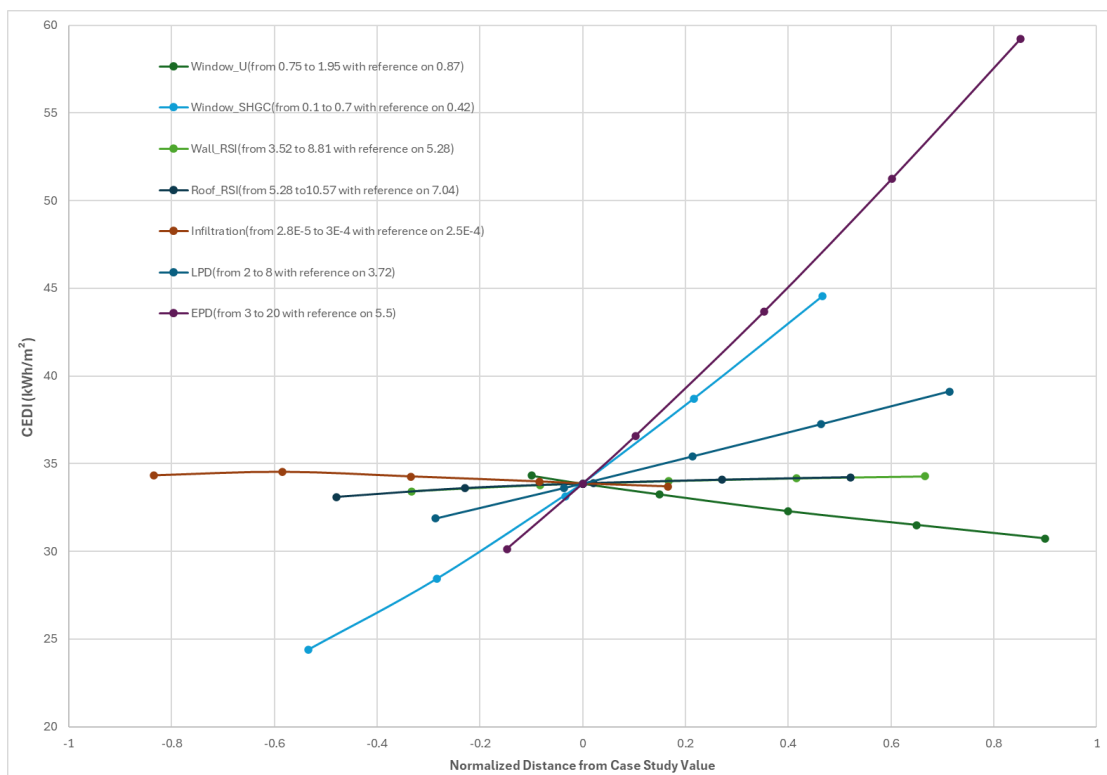


Figure 5-19- Design parameter sensitivity analysis on CEDI

Chapter 6

6 Results and Analysis

6.1 Introduction

This chapter presents the results of the surrogate modelling framework developed for predicting both hourly and annual energy performance of high-performance office buildings in Waterloo. The results are organized to reflect the logical sequence of the modelling process, including data generation, preprocessing, model training and tuning, ensemble prediction testing, and sensitivity analysis. The GitHub link for accessing the Python implementation of the surrogate model is provided in “Appendix-C: Python”. Versions of Python and all libraries used in this study is presented in “Appendix-B: Python Code”.

Each section corresponds to the numbered steps presented under Chapter 3 Methodology, which were systematically executed to ensure transparency and reproducibility of the surrogate modelling process.

The analytical objectives of this chapter are threefold:

1. To evaluate the accuracy and robustness of three surrogate models developed using fully connected artificial neural networks (FC-ANN; MLP), convolutional neural networks (CNN) and an ensemble combining both architectures.
2. To assess the surrogate model’s ability to reproduce hourly energy consumption profiles and key annual indicators under a Typical Meteorological Weather (TMY) for the City of Waterloo.
3. To identify the most influential input parameters on energy consumption through comprehensive sensitivity analysis.

6.2 Dataset Generation and Preprocessing

The data generation pipeline produced a synthetic dataset comprising 3,000 simulations, each corresponding to a unique combination of design parameters. This dataset was created using a custom-developed Python–EnergyPlus API script, which automated the modification of input data files (IDF) and extraction of hourly and annual simulation outputs.

6.2.1 Dataset Overview

The generated dataset represents variations in twelve input parameters, categorized as *geometry-related* (Area, Aspect Ratio, WWR) and *simple parameters* (e.g., window properties, thermal resistance,

infiltration, and internal load intensities). The full parameter set, variation ranges, and type classifications are summarized in Table 5-1 and Table 5-2.

Two major types of outputs have been extracted from each simulation:

1. **Hourly energy consumption profile:** a matrix of 8,760 values (365 days $\times$ 24 hours) representing total hourly building electricity use for a full calendar year.
2. **Annual performance indicators:** Energy Use Intensity (EUI), Thermal Energy Demand Intensity (TEDI), and Cooling Energy Demand Intensity (CEDI), expressed in kWh/m² of conditioned floor area.

The generation of synthetic data ensured a representative coverage of the input-parameter space by leveraging LHS method, covering both typical and extreme operational scenarios.

6.2.2 Preprocessing and Feature Scaling

Before training, input features were normalized using Min–Max scaling to confine all parameters within the range of [0,1]. This scaling was essential for stable ANN and CNN convergence, as the raw parameter magnitudes differed by several orders (e.g., infiltration $\sim 10^{-4}$ versus Building Footprint Area $\sim 10^3$).

Outlier detection was carried out using the IQR method to ensure physical plausibility and to remove corrupted simulation results. No outliers were detected, confirming the numerical stability and physical consistency of the parametric model across the full parameter space, consistent with the verification results presented in Chapter 5.

The final cleaned dataset contained 3000 valid simulations, which were then split into training (85%) and final testing (15%) subsets. Within the training subset, a five-fold cross-validation scheme was employed for robust model tuning.

6.3 Model Selection and Structure

The selection of surrogate modelling approaches in this study is guided by the need to predict high-dimensional, time-resolved building energy outputs from a relatively low-dimensional set of design parameters. This presents a modelling challenge that extends beyond conventional regression or scalar prediction tasks, requiring approaches capable of capturing both nonlinear relationships and structured temporal dependencies.

Artificial Neural Networks (ANNs), specifically multilayer perceptron (MLP), is employed due to its established capability in modelling complex nonlinear interactions between input variables and building

energy performance. In comparison to parametric regression models, ANNs provide greater flexibility in representing relationships involving occupancy behavior, HVAC operation, and environmental conditions. Within this study, fully connected neural network (FC-ANN or MLP) models serve as a baseline non-parametric approach, enabling comparison with simpler regression methods while maintaining computational efficiency. However, FC-ANN models do not explicitly capture structured temporal dependencies in high-dimensional outputs such as full-year hourly energy profiles.

To address the challenge of predicting full-year hourly energy profiles, the problem is reformulated as a structured output prediction task. Instead of treating energy prediction as a sequence of independent time steps, the hourly energy profile is represented as a two-dimensional structure (day $\times$ hour). This representation allows the use of Convolutional Neural Networks (CNNs), which are designed to extract spatial patterns and local dependencies in structured data. By leveraging this formulation, the model is able to capture recurring temporal features such as daily cycles, seasonal transitions, and operational patterns more effectively than conventional point-wise prediction approaches. This formulation also enables the model to implicitly capture temporal dependencies (autocorrelation) in energy consumption patterns, without requiring explicit sequential modelling.

This two-dimensional representation leverages the inherent structure of building energy data, enabling the model to learn both the magnitude and shape of energy profiles rather than treating each time step as an independent prediction target.

In addition, a hybrid ANN–CNN architecture is explored to combine the strengths of both modelling approaches. While MLP components effectively process structured input parameters, CNN components are well-suited for generating structured, high-dimensional outputs. The integration of these components allows the model to simultaneously capture nonlinear input–output relationships and structured temporal dependencies within a unified framework.

6.3.1 Baseline Architectures: FC-ANN and CNN

Two types of surrogate models were explored in this first step:

- **Fully Connected Artificial Neural Networks (FC-ANN):** Fully connected feedforward networks designed to approximate nonlinear mappings between the input features and the target outputs (annual and hourly). Prior studies have reported strong performance of ANNs for predicting aggregate outcomes such as annual energy consumption [57], [84]. Therefore, FC-ANNs were included as one of the baseline models in the model selection process.

- **Convolutional Neural Networks (CNN):** Convolutional Neural Networks, designed to extract spatial and structural patterns from two-dimensional data [29], [85], are well-suited to hourly energy profiles represented as 2D matrices, enabling detection of recurring daily cycles and seasonal transitions.

Both baseline models were developed in TensorFlow, with Optuna employed for automated hyperparameter optimization.

6.3.2 Hyperparameter Optimization Process

A systematic hyperparameter optimization process was performed using Optuna [134], aiming to identify the most effective model structure and configuration for predicting both annual energy indicators (EUI, TEDI, CEDI) and the hourly load profiles. Each architecture was optimized using 150 Optuna trials under a 5-fold cross-validation regime. The optimization objective was to minimize the total composite loss defined as:

$$\text{Total Loss} = \text{MSE}_{\text{annual}} + \alpha \times \text{MSE}_{\text{hourly}} \quad \text{Eq 6-1}$$

where $\alpha = w_{\text{hour}}$ is the weighting factor that balances annual and hourly prediction accuracy.

The models were trained using the Adam optimizer with early stopping based on validation loss convergence. Optuna's Tree-structured Parzen Estimator (TPE) algorithm adaptively explored the high-dimensional hyperparameter space, focusing on promising configurations over successive trials.

6.3.2.1 Hyperparameter Optimization for ANN

For the fully connected ANN, Optuna executed 150 trials, exploring a range of hyperparameters, as shown in Table 3-1. The hyperparameter optimal values are presented in Table 6-1.

Table 6-1- Optimized hyperparameter values, using Optuna for FC-ANN

Hyperparameter	Optimized Value using Optuna
latent_dim	256
hidden_dim	256
lr	8.013×10^{-3}
wd	3.272×10^{-6}
w_hour (α)	0.165
Minimized Total Composite Loss Value	3.08×10^{-4}

The relatively low optimal w_{hour} indicates that the ANN's dense layers performed effectively in predicting aggregate (annual) energy metrics but struggled to model detailed hourly variations. Increasing

w_{hour} caused overfitting to noisy short-term patterns without improving generalization, suggesting the ANN alone lacked the structural capacity to model temporal dependencies.

6.3.2.2 Hyperparameter Optimization for CNN

Similarly, for the CNN, Optuna executed 150 trials. The hyperparameter optimal values are presented in Table 6-2.

Table 6-2 Optimized hyperparameter values, using Optuna for CNN

Hyperparameter	Optimized Value based using Optuna
Latent dimension	256
Convolutional channels	16
CNN channel multiplier	128
Learning rate	1.185×10^{-3}
Weight decay	1.611×10^{-6}
Hourly loss weighting (w_{hour} or α)	1.009
Use of 2d CNN	True
Minimized Total Composite Loss Value	1.499×10^{-3}

The higher optimal w_{hour} value indicates that, within the adopted loss formulation and Optuna search space, the best-performing CNN configuration required a greater relative weighting on the hourly loss term to minimize the overall objective. While this adjustment improved the model's ability to reduce hourly prediction error, it was accompanied by slower convergence and a substantially higher computational cost compared to the ANN-based approach.

Although the CNN architecture demonstrated improved performance for hourly predictions, based on the lower hourly losses for this model, its optimal Optuna configuration resulted in a higher total validation loss compared to the ANN model. This indicates that, despite stronger temporal feature extraction, the CNN's overall loss balance was less favorable at the hyperparameter optimization stage. Therefore, further evaluation during the full training and validation phase was required to assess whether loss-weight optimization could improve the CNN's total performance.

6.3.3 Model Training and Validation of Baseline Architectures

The optimized ANN and CNN models were retrained on the complete 85% training dataset using their respective optimal hyperparameters. A 5-fold cross-validation approach was employed to enhance model generalization and robustness. In each fold, an independent model was trained and validated, and the optimized weights were recorded. The final ensemble model was constructed by aggregating the

predictions of the individual fold models (via averaging) during inference. This ensemble approach reduces variance and improves predictive stability without requiring explicit weight averaging. The ensemble performance was then evaluated on the unseen 15% test dataset.

For each model (ANN and CNN) total validation losses during train process have been presented in Figure 6-1, and Figure 6-2, respectively. It should be noted that, due to the relatively high loss values during the initial training epochs of the CNN model, it becomes difficult to observe the range of loss variations in the later stages of training. To address this, a zoomed-in plot of the CNN validation losses is provided to better illustrate the model's performance during the later epochs. Both Figure 6-1, and Figure 6-2, illustrate smooth convergence of total loss for FC-ANN and CNN models for all 5 folds.

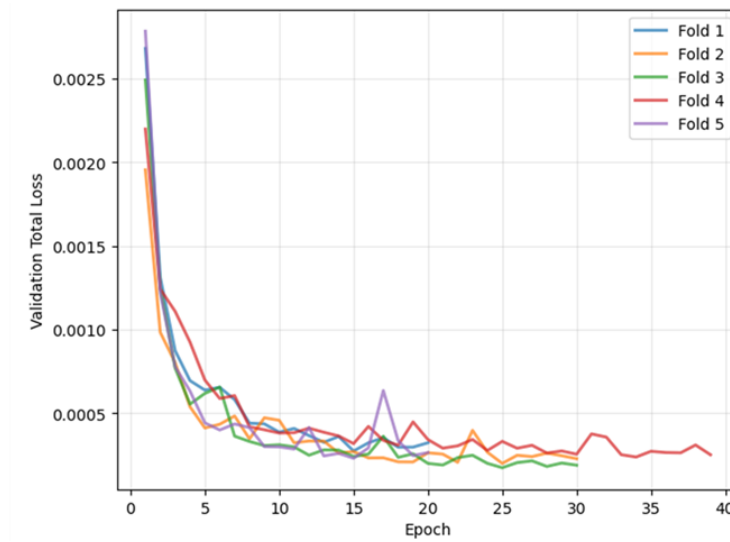


Figure 6-1- Total validation losses for fully connected ANN model during training process for all folds

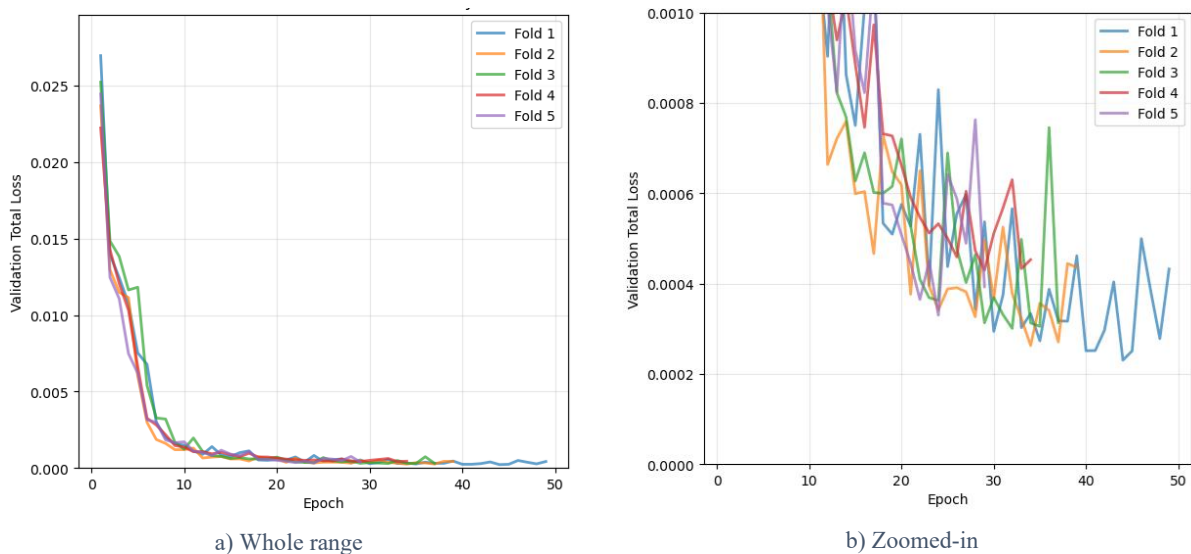


Figure 6-2- Total validation losses for CNN model during training process for all folds

One of the key evaluation aspects for the developed models is their comparative performance on annual and hourly prediction tasks. To assess this, both annual and hourly validation losses, along with their ratio (Hourly / Annual), were analyzed for each model. This ratio provides an interpretable measure of whether a model prioritizes long-term aggregate behavior (annual indicators) or short-term temporal dynamics (hourly profiles).

For the ANN model, the validation losses for annual and hourly outputs, as well as the corresponding hourly-to-annual loss ratios across all five folds during training, are illustrated in Figure 6-3. Similarly, Figure 6-4 presents the same validation loss and ratio trends for the CNN model.

As illustrated in Figure 6-3(a) and (b), the fully connected ANN exhibits a rapid reduction in both annual and hourly validation losses during the initial training phase, with the most pronounced decrease occurring within the first 5 epochs across all folds. After this initial convergence phase, the losses decrease more gradually and eventually stabilize, indicating effective learning and well-controlled optimization. Based on Figure 6-3(c), throughout training, the validation losses for annual outputs remain mostly lower than those for hourly predictions, resulting in hourly-to-annual loss ratios predominantly greater than one. As shown in Figure 6-3(d), the best losses obtained for all folds indicate that hourly prediction errors remain higher than annual errors, reflecting the greater complexity of learning high-resolution temporal outputs. This pattern reflects the ANN's stronger performance in learning low-dimensional, aggregated annual energy indicators compared to high-dimensional hourly profiles, as expected. Within the adopted loss formulation, the relatively low optimal w_{hour} value further limited the contribution of the hourly term, yielding stable and comparatively low total validation losses.

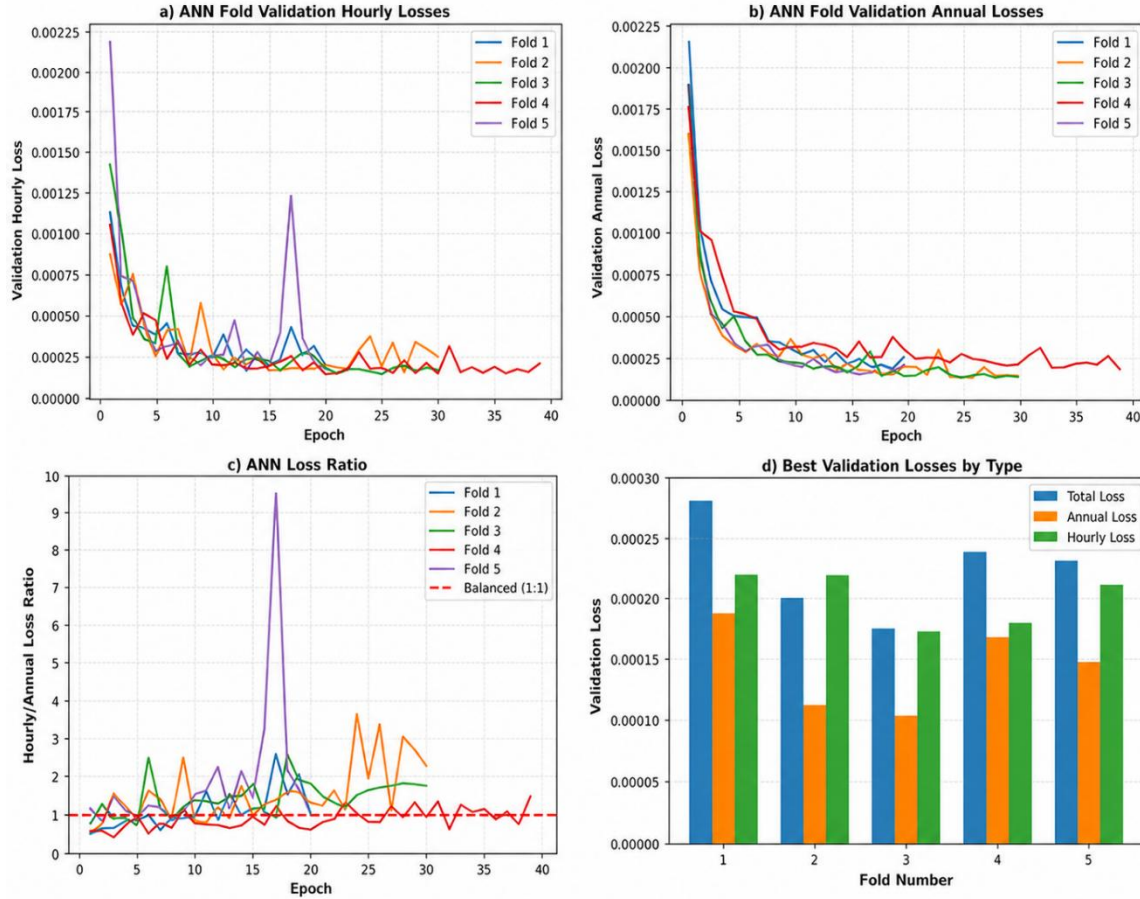


Figure 6-3- Individual validation losses and loss ratio for fully connected ANN model.

Figure 6-4(a) and (b), show that the CNN also achieves a sharp reduction in validation losses during the early epochs (first 10 epochs), followed by a smoother convergence phase similar to that observed for the ANN. However, the relative behavior of the loss components differs substantially. The CNN attains lower validation losses for hourly outputs than for annual metrics, with the hourly-to-annual loss ratio remaining consistently below one and, in several folds, dropping below 0.5, as shown in the Figure 6-4(c). This indicates that, under the selected loss formulation and optimization settings, emphasizing the hourly component was beneficial for CNN training. Furthermore, based on Figure 6-4(d) for all folds, except for fold-2, the best hourly losses are substantially less than the best annual losses. Nevertheless, the higher annual validation losses observed for the CNN exert a stronger influence on the composite objective, leading to a higher total validation loss compared to the ANN. As a result, despite its strong capability in modelling high-resolution hourly patterns, the CNN demonstrates lower overall performance when evaluated using the combined loss metric.

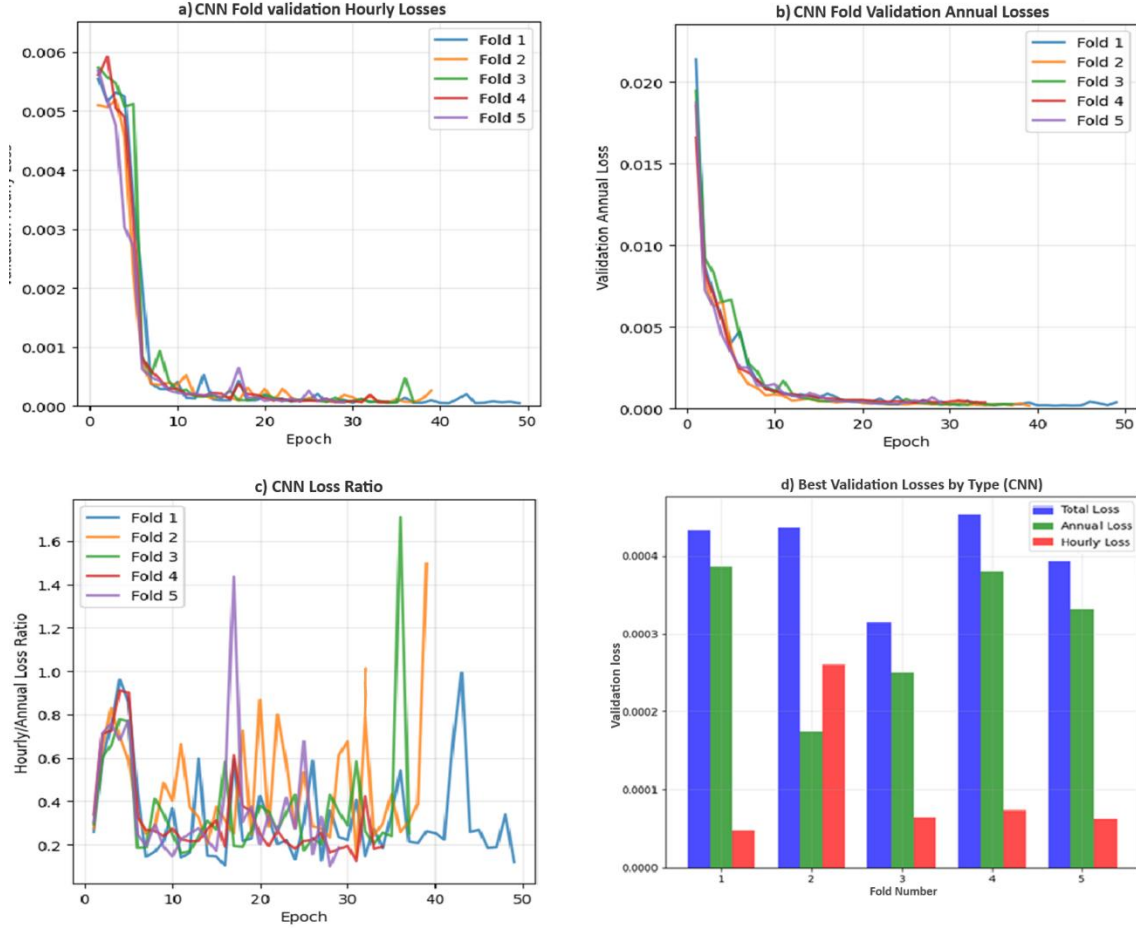


Figure 6-4- Individual validation losses and loss ratio for CNN model

The ensemble performance summaries for both the fully connected (FC) ANN and CNN models, presented in Table 6-3 and Table 6-4 respectively, provide quantitative insight into the consistency and accuracy of each model across the five validation folds. The ANN model demonstrated stable convergence, with a mean total loss of 0.000223 ± 0.000034 . Its mean annual and hourly losses were 0.000143 ± 0.000031 and 0.000200 ± 0.000020 , respectively, corresponding to a mean hourly-to-annual loss ratio of approximately 1.40. The low standard deviation values indicate that the ANN maintained reliable generalization and repeatability across all folds.

In contrast, the CNN model exhibited a lower mean hourly loss (0.000099 ± 0.000008), confirming its capacity to capture fine-grained temporal variations. However, this improvement was accompanied by higher mean total (0.000394 ± 0.000068) and annual (0.000303 ± 0.000063) losses, with a much lower hourly-to-annual loss ratio of approximately 0.327. The CNN also showed greater variability between folds and required a higher number of training epochs (24–44) for convergence, reflecting its greater sensitivity to initialization and hyperparameter configuration. It should be noted that each epoch of CNN model demanded around twice the runtime of the FC-ANN under identical hardware and batch size.

Table 6-3- ANN Ensemble model summary

Fold number	Best total loss	Best annual loss	Best hourly loss	Loss ratio (hourly per annual)	Best epoch
1	2.74×10^{-4}	1.86×10^{-4}	2.19×10^{-4}	1.18	15
2	2.00×10^{-4}	1.13×10^{-4}	2.18×10^{-4}	1.93	25
3	1.74×10^{-4}	1.05×10^{-4}	1.72×10^{-4}	1.64	25
4	2.38×10^{-4}	1.66×10^{-4}	1.79×10^{-4}	1.08	34
5	2.30×10^{-4}	1.46×10^{-4}	2.10×10^{-4}	1.44	15
Ensemble Mean (standard deviation)	2.23×10^{-4} (3.4×10^{-5})	1.43×10^{-4} (3.1×10^{-5})	2.00×10^{-4} (2.0×10^{-5})	1.45	

Table 6-4- CNN Ensemble model summary

Fold number	Best total loss	Best annual loss	Best hourly loss	Loss ratio (hourly per annual)	Best epoch
1	4.21×10^{-4}	3.83×10^{-4}	4.7×10^{-5}	0.123	44
2	4.20×10^{-4}	1.86×10^{-4}	2.51×10^{-4}	1.360	34
3	3.10×10^{-4}	2.38×10^{-4}	6.3×10^{-5}	0.265	32
4	4.40×10^{-4}	3.73×10^{-4}	6.6×10^{-5}	0.177	29
5	3.93×10^{-4}	3.36×10^{-4}	7.1×10^{-5}	0.211	24
Ensemble Mean (standard deviation)	3.94×10^{-4} (6.8×10^{-5})	3.03×10^{-4} (6.3×10^{-5})	9.9×10^{-5} (8×10^{-6})	0.327	

6.3.4 Model Performance on the Held-Out Test Set

After completing the 5-fold cross-validation and final training, the optimized ANN and CNN models were evaluated on the held-out 15% test dataset, which was not used for hyperparameter tuning or training. This independent evaluation allows for a fair assessment of each model's generalization capability and predictive performance on unseen building configurations.

6.3.4.1 ANN Model Performance on Test Set

The ANN demonstrated strong predictive performance on annual indicators, achieved low errors for EUI, TEDI, and CEDI predictions across the test set but its performance degraded when evaluated on the hourly energy consumption profiles.

Key quantitative metrics for ANN performance on the unseen test dataset are summarized in Table 6-5. The quality of the fitness of predicted annual and hourly data outputs versus true values are shown in Figure 6-5. The fully connected ANN exhibits excellent predictive accuracy for the annual performance indicators (EUI, TEDI, and CEDI), achieving R^2 values above 0.99 and RMSE values below 1.5 kWh/m². These results confirm the network’s capability to capture annual energy behavior. However, the model’s performance declines when evaluated on hourly energy data, as reflected by the higher RMSE (≈ 14.9 kWh). This indicates that while the ANN effectively learns aggregated relationships between input parameters and overall energy consumption, it struggles to accurately reproduce short-term temporal dynamics, particularly the rapid fluctuations associated with cooling loads and equipment operation.

Table 6-5- Evaluation metrics for fully connected ANN performance on the unseen test dataset

Output metric	RMSE	MAE	R²
EUI	1.00 kWh/m ²	0.76 kWh/m ²	0.998
TEDI	0.75 kWh/m ²	0.59 kWh/m ²	0.993
CEDI	1.32 kWh/m ²	0.94 kWh/m ²	0.996
Hourly	14.94 kWh	9.82 kWh	0.984

The performance of the FC-ANN model can be further examined using the surrogate-versus-numerical³ value distributions on the held-out test dataset, as shown in Figure 6-5. In Figure 6-5(a) to (c), the predicted values for the annual energy indices are generally concentrated around the unity line, and the fitted regression lines closely follow this reference. This indicates that the FC-ANN captures the overall trends and magnitudes of the annual metrics with reasonable accuracy across the evaluated range.

In contrast Figure 6-5 (d) reveals noticeable discrepancies between the predicted and true distributions for hourly energy consumption. The misalignment between the red (predicted) and blue (true) density distributions in some columns indicates larger errors and reduced fidelity in reproducing detailed hourly patterns. These deviations highlight a limitation of the FC-ANN in modelling high-resolution temporal outputs, confirming that the high-dimensional temporal structure of hourly profiles requires an architecture with explicit spatial pattern recognition capability, motivating the CNN baseline evaluated in the following section.

³ In this study, the terms *numerical* and *surrogate* are used instead of *true* and *predicted*, where *numerical* refers to outputs generated by the physics-based EnergyPlus simulations.

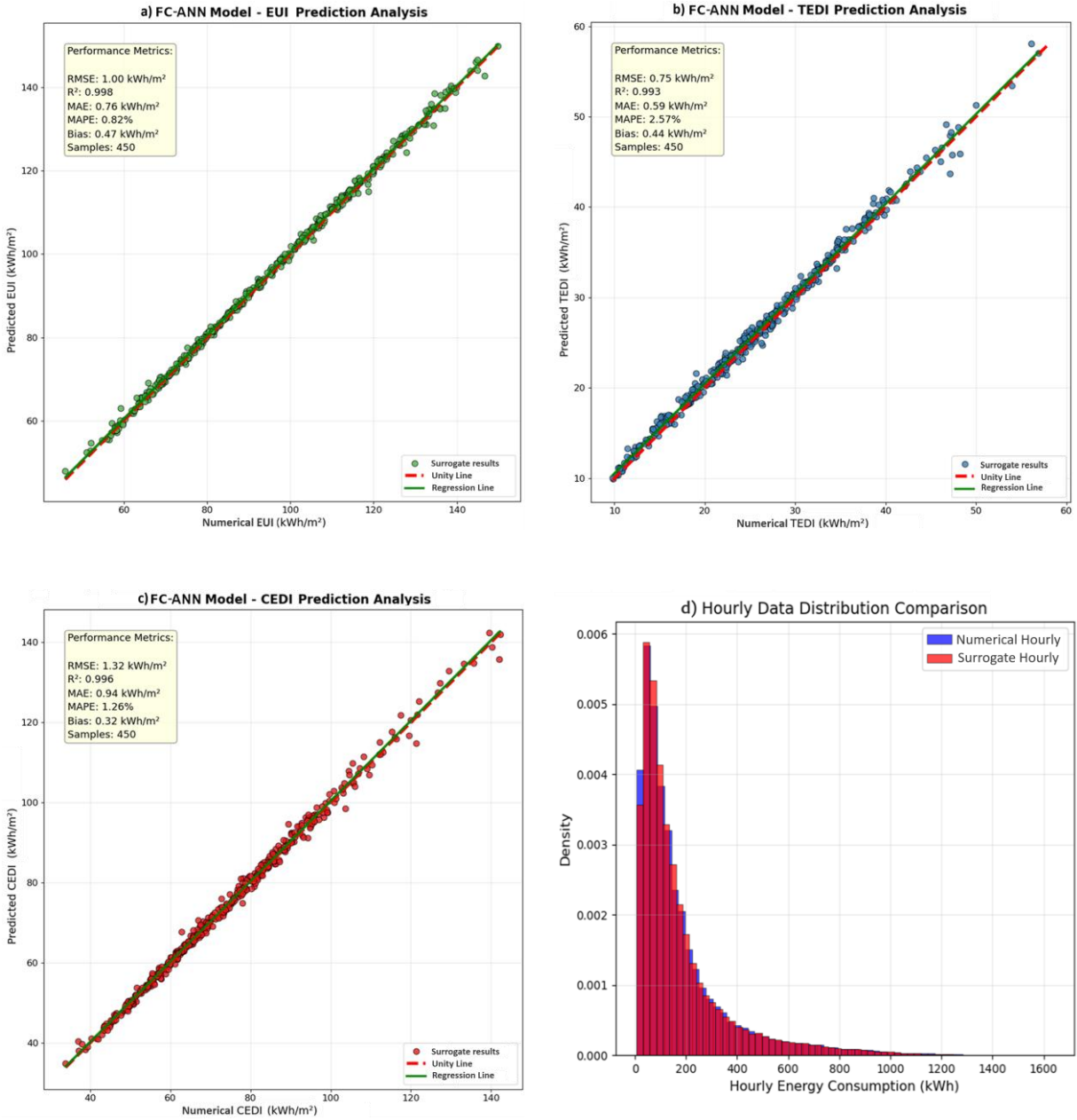


Figure 6-5- Numerical versus Surrogate for fully connected (FC) ANN model evaluation on test dataset

6.3.4.2 CNN Model Performance on Test Set

The CNN, which leverages convolutional layers to detect temporal and structural dependencies, achieved higher accuracy for hourly energy consumption compared to the ANN. The convolutional filters enabled CNN to recognize repeating daily patterns and transient load variations more effectively. However, CNN training was computationally more demanding, with longer convergence times and occasional higher losses in annual indicator predictions.

Table 6-6 presents the quantitative metrics evaluating the performance of the CNN on the unseen test dataset while Figure 6-6 illustrates the accuracy of the predicted annual and hourly outputs in comparison to the actual values.

Table 6-6- Evaluation metrics for CNN model performance on the unseen test dataset

Output metric	RMSE	MAE	R²
EUI	1.06 kWh/m ²	0.77 kWh/m ²	0.997
TEDI	0.99 kWh/m ²	0.75 kWh/m ²	0.988
CEDI	1.86 kWh/m ²	1.32 kWh/m ²	0.990
Hourly	12.664 kWh	8.746 kWh	0.996

As observed from Table 6-6 and Figure 6-6, CNN demonstrates notable improvements in capturing hourly energy consumption dynamics, outperforming the ANN in both RMSE and R² for hourly predictions (RMSE = 12.66 kWh, R² = 0.996).

However, the CNN's performance on the prediction of the annual indicators (EUI, TEDI, and CEDI) was slightly lower than the ANN. A plausible explanation is that the model inputs are 12 scalar (tabular) parameters, which provide limited local structure for convolutional kernels to exploit at the input stage. Consequently, the CNN's comparative advantage emerges primarily in modelling the hourly outputs, which inherently has a 2D temporal structure (365×24), rather than in learning richer representations from the tabular feature set. This realization motivated the development of a hybrid ANN–CNN architecture, designed to leverage the ANN's strength in handling tabular inputs while utilizing convolutional layers to model the temporal dependencies of the hourly output domain.

The predictive performance of the CNN model on the held-out test dataset is illustrated in Figure 6-6. As shown in Figure 6-6(a), the surrogate predictions for EUI are closely aligned with the unity line, and the fitted regression line follows the ideal trend across most of the value range, indicating reliable performance for this aggregate metric. In contrast, Figure 6-6(b) and (c), corresponding to TEDI and CEDI, exhibit noticeably larger dispersion around the unity line. For these indices, the regression lines deviate slightly from the ideal slope, particularly at higher values, suggesting reduced predictive accuracy for extreme heating- and cooling-dominated cases.

By comparison, the hourly energy distribution presented in Figure 6-6(d) demonstrates a substantially improved agreement between numerical and surrogate-predicted data. The predicted (red) density closely matches the numerical (blue) distribution across most bins, indicating that the CNN effectively captures the temporal structure and variability of hourly energy consumption. This result is consistent with the

lower hourly validation losses observed during training and confirms the CNN's relative strength in high-resolution, time-dependent energy prediction.

Furthermore, a comparison of the hourly energy density distributions shown in Figure 6-5(d) and Figure 6-6(d), reveals that the surrogate distribution (red) produced by the CNN aligns more closely with the Numerical distribution (blue) than that of the ANN. This improved overlap on the unseen test dataset further confirms the CNN's higher capability in capturing the statistical characteristics of hourly energy consumption patterns.

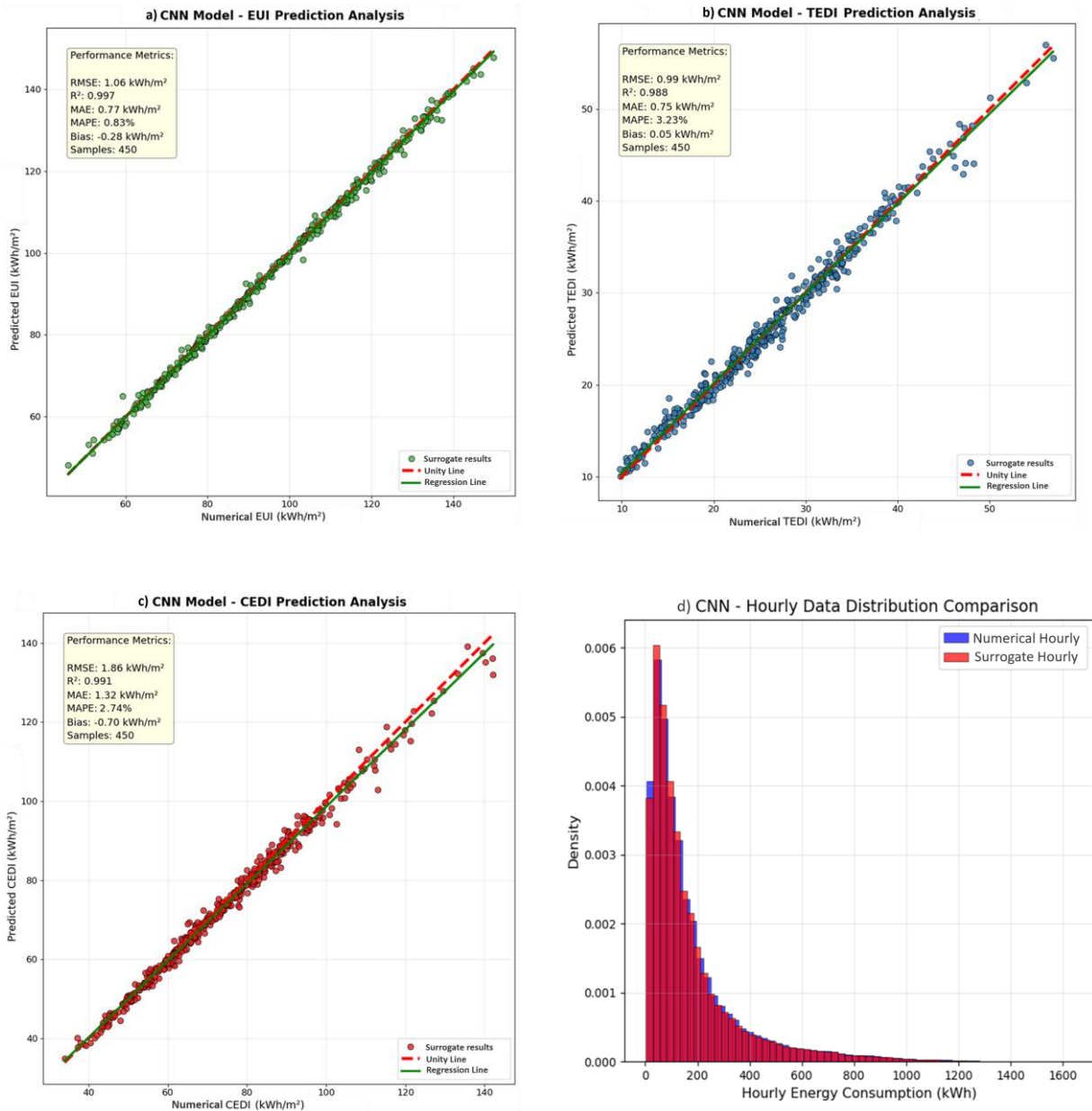


Figure 6-6- Numerical versus Surrogate value for CNN model evaluation on test dataset

6.4 Hybrid ANN–CNN Model

6.4.1 Model Architecture (ANN with Convolutional Output Head)

To address the complementary strengths and weaknesses of the standalone ANN and CNN models, a hybrid ANN–CNN architecture was developed, beyond the baseline models. This approach integrates the global parametric learning capability of an ANN with the temporal feature extraction power of a convolutional head attached to the model outputs. The design specifically targets the mismatch between the tabular input space (12 independent features) and the structured hourly output space (365×24), ensuring that each component of the network operates on the data type it models best.

6.4.1.1 Architectural Concept

In this hybrid structure, the fully connected ANN acts as the feature extraction backbone, transforming the 12-dimensional input vector into a latent representation that encapsulates overall building energy behavior. Instead of connecting this latent representation directly to dense output neurons, it is reshaped and passed through a convolutional output head that reconstructs the hourly energy profile. This structure enables the model to implicitly capture and reconstruct temporal dependencies in energy consumption patterns, including daily and seasonal autocorrelation effects identified in the data.

This hybrid connection enables the model to simultaneously capture global relationships between design parameters and overall energy performance, as well as fine-grained temporal variations within the hourly energy profile.

6.4.1.2 Model Structure

The hybrid ANN–CNN architecture is systematically designed based on the complementary strengths of the baseline ANN and CNN models. The ANN demonstrated strong performance on low-dimensional tabular inputs and annual metrics, while the CNN excelled at capturing temporal structure in the hourly output domain but struggled with scalar feature encoding and computational efficiency. The macro-architecture was therefore fixed to enforce an appropriate inductive bias, while Optuna was used to optimize model capacity and training-related hyperparameters within predefined bounds. In practice, the optimization process was restricted to parameters such as the number of hidden units, learning rate, dropout rate, and convolutional filter sizes, while the overall ANN–CNN structure remained unchanged.

Figure 6-7 illustrates the hybrid ANN–CNN architecture and the flow of information from tabular inputs to multi-scale outputs. This hybrid ANN–CNN architecture aims to resolve the input–output structure mismatch as follows:

- **ANN backbone:** optimized for low-dimensional tabular inputs.

- **CNN head:** optimized for high-dimensional temporal outputs.
- **Joint training:** enhances temporal fidelity while maintaining annual-metric accuracy.

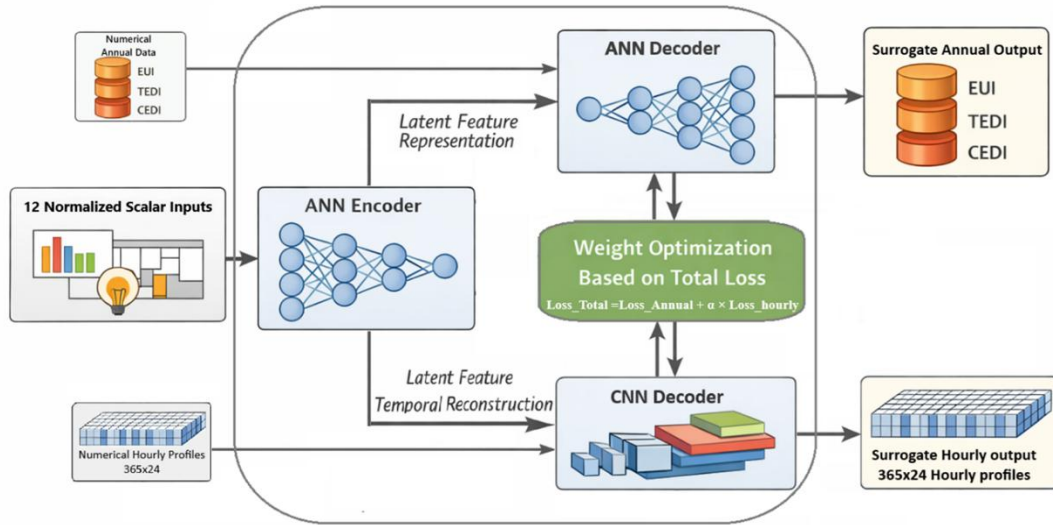


Figure 6-7- Schematic of hybrid model architecture

1. ANN Encoder (Feature Extraction Backbone)

- Input: 12 normalized scalar parameters (e.g., Area, Aspect Ratio, LPD, EPD).
- Layers: Four fully connected layers (optimized by Optuna) with ReLU activation (as it is mentioned as regular activation function used in the literatures) and dropout (0.2).
- Output: A latent feature vector summarizing the building's energy behavior.
- Role: Captures high-level relationships between design and operational parameters.

2. Convolutional Output Head

- Input: Latent vector reshaped into pseudo-temporal format (e.g., [batch, channels, time]).
- Layers: Stack of 1D convolutional layers (With optimized channels size, and kernel sizes 3–5–7 based on standard used in the literature) with ReLU activation.
- Output: Full 365×24 hourly energy consumption sequence.
- Purpose: Learns cyclic temporal dependencies and localized energy variation patterns without adding excessive computational cost.

3. Dual Output Branches

- Annual Output Branch: A dense layer maps the latent features directly to annual indicators (EUI, TEDI, CEDI).
- Hourly Output Branch: The convolutional head reconstructs the hourly consumption profile.
- This multi-output design encourages consistency between aggregated and time-resolved predictions.

6.4.2 Hyperparameter Optimization for the Hybrid ANN–CNN Model

Same as the data split method that mentioned for ANN and CNN models, each trial trained the model on 85% of the dataset, reserving 15% unseen data for final test. The hyperparameter optimization for the hybrid ANN–CNN model was conducted using the Optuna framework, employing a Bayesian optimization approach to systematically explore the parameter space and minimize the composite loss function in Eq 6-1. Optuna trials were executed on unique combination of hyperparameters sampled within the mentioned ranges and the best selected value presented in Table 6-7.

Table 6-7- Optimized hyperparameter values, using Optuna for Hybrid model

Hyperparameter	Optimized Value using Optuna
latent_dim	128
conv_channels	16
lr	4.623×10^{-3}
wd	1.732×10^{-6}
w_hour (α)	0.316
Optimizer	Adam
Batch size	64
Minimized Total Composite Loss Value	2.71×10^{-4}

The objective function minimized the validation total loss (2.71×10^{-4}), computed as the mean overall output branches and temporal dimensions. The moderate latent dimension (128) provided a balance between model expressiveness and generalization, while the 16 convolutional channels allowed the output head to effectively capture dominant temporal patterns without overfitting. The hourly weighting coefficient ($\alpha \sim 0.32$) indicates that the model allocated roughly one-third of its learning focus to refining

the hourly temporal fidelity, aligning well with the intended trade-off between temporal detail and annual stability.

6.4.3 Training and Validation of the Hybrid Model

Following the hyperparameter optimization process, the hybrid ANN–CNN model was retrained using the full 85% training dataset and validated through a 5-fold cross-validation procedure. Each fold independently optimized its model weights to minimize the composite loss function, balancing the annual and hourly objectives as defined earlier.

Figure 6-8 illustrates the total validation loss trajectories for all five cross-validation folds during training of the hybrid ANN–CNN model. Each curve corresponds to one-fold and represents the composite validation loss aggregated across all output branches and temporal dimensions. Across all folds, the total validation loss decreases rapidly during the early epochs, indicating efficient learning and stable optimization behavior. The curves then gradually plateau, suggesting convergence of the model and effective early stopping without signs of overfitting. The final loss values for the different folds lie within a narrow range, demonstrating low variability across data splits and consistent generalization performance.

The similarity in convergence patterns and terminal loss magnitudes across folds indicates that the hybrid architecture is robust to variations in the training–validation partitioning. In addition, the loss trajectories closely resemble those observed for the fully connected ANN baseline, suggesting that the ANN backbone preserves training stability while the convolutional output head does not introduce instability into the optimization process.

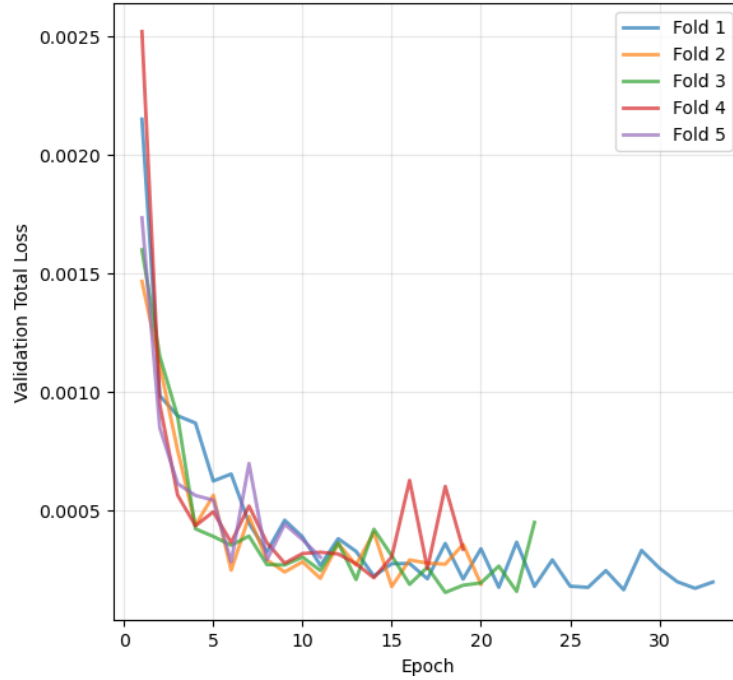


Figure 6-8- Total validation losses for each fold during training process for Hybrid ANN-CNN model

Overall, Figure 6-8 confirms that the hybrid ANN–CNN model achieves stable convergence, fold-to-fold consistency, and reliable total loss minimization, providing a sound basis for further analysis of its individual annual and hourly loss components presented in Figure 6-9.

Figure 6-9 presents the evolution of the individual validation loss components for the hybrid ANN–CNN model across all five cross-validation folds, including hourly loss in Figure 6-9(a), annual loss in Figure 6-9(b), the hourly-to-annual loss ratio in Figure 6-9(c), and the best validation loss in Figure 6-9(d). For both hourly and annual outputs, all folds exhibit a sharp reduction in validation loss during the initial training epochs, followed by a gradual stabilization, indicating efficient learning and well-controlled optimization. Early stopping was consistently triggered after convergence, preventing overfitting and ensuring stable generalization behavior.

The best validation losses achieved for both hourly and annual outputs lie within a narrow range across folds, demonstrating low variability and consistent performance regardless of data partitioning. This confirms that the hybrid architecture maintains balanced predictive accuracy for both scalar annual metrics and high-dimensional hourly profiles.

The hourly-to-annual loss ratio further highlights the effectiveness of the hybrid design. Across folds, the average ratio remains approximately 0.6, indicating a well-balanced contribution of hourly and annual losses to the total objective. This ratio lies between that of the fully connected ANN model (≈ 1.5), which favored annual predictions, and the CNN model (≈ 0.3), which emphasized hourly accuracy at the expense

of annual performance. The hybrid model therefore achieves a compromise that preserves strong hourly fidelity while maintaining competitive annual accuracy.

Overall, Figure 6-9 demonstrates that the hybrid ANN–CNN architecture successfully balances multi-scale prediction objectives, yielding stable convergence, fold-to-fold consistency, and improved loss trade-offs relative to the baseline ANN and CNN models.

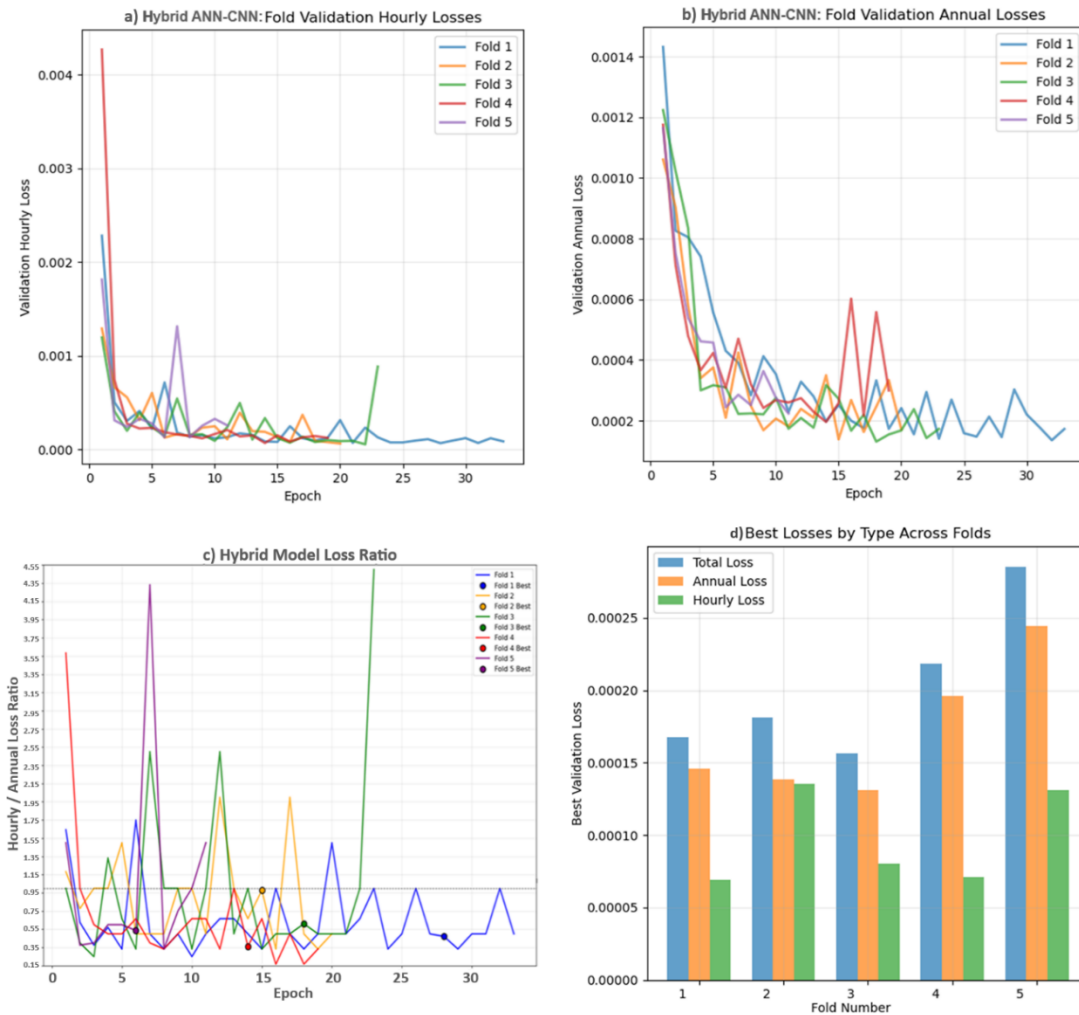


Figure 6-9- Individual validation losses and loss ratio for hybrid ANN-CNN model

The final hybrid model implemented as an ensemble ANN-CNN architecture that combines the strengths of fully connected and convolutional representations, achieved validation results summarized in Table 6-8. A direct comparison with the fully connected ANN (Table 6-3), and CNN (Table 6-4) highlights the hybrid model’s balanced performance across both annual and hourly prediction scales.

Table 6-8- Ensemble validation losses for hybrid ANN-CNN model.

Fold number	Best total loss	Best annual loss	Best hourly loss	Loss ratio (hourly per annual)	Best epoch
1	1.64×10^{-4}	1.46×10^{-4}	6.9×10^{-5}	0.47	28
2	1.81×10^{-4}	1.38×10^{-4}	1.35×10^{-4}	0.98	15
3	1.56×10^{-4}	1.31×10^{-4}	8.0×10^{-5}	0.61	18
4	2.18×10^{-4}	1.96×10^{-4}	7.1×10^{-5}	0.36	14
5	2.85×10^{-4}	2.44×10^{-4}	1.31×10^{-4}	0.54	6
Ensemble Mean (standard deviation)	2.02×10^{-4} (4.7×10^{-5})	1.70×10^{-4} (4.3×10^{-5})	9.7×10^{-5} (3.0×10^{-5})	0.58	

Numerically, the ensemble mean total loss of the hybrid model (2.02×10^{-4}) is lower than that of both the ANN (2.23×10^{-4}) and CNN (3.11×10^{-4}) models, demonstrating improved overall generalization. The annual loss (1.71×10^{-4}) is slightly higher than that of the ANN (1.43×10^{-4}) but remains significantly lower than the CNN (2.49×10^{-4}), while the hourly loss (9.7×10^{-5}) represents a substantial improvement over the ANN (2.0×10^{-4}) and is reasonably close to the CNN (6.1×10^{-5}). This indicates that the hybrid model successfully bridges the gap between the global accuracy of the ANN and the temporal sensitivity of the CNN.

Examining fold-by-fold performance, the hybrid architecture consistently achieved low total losses across all training folds, with early stopping in Fold 5 explaining its slightly higher loss value. Importantly, the ensemble total loss, computed as the aggregated performance of all folds, remained the lowest among all three architectures, confirming the hybrid model's superior generalization stability.

From a computational standpoint, the hybrid approach required moderately more training time than the ANN but substantially less than the CNN, offering a favorable trade-off between performance and efficiency. The results suggest that the hybrid ANN-CNN model leverages the ANN's capability to model global relationships in non-spatial input data (e.g., 12 scalar parameters) while benefiting from the CNN head's ability to extract temporal structures from the 2D hourly output space.

6.4.4 Model Performance on the Held-Out Test Set

When evaluated on the unseen 15% test dataset, the hybrid ANN-CNN model demonstrated the best overall generalization performance, among evaluated models, successfully balancing the accuracy of the fully connected ANN and the temporal sensitivity of the CNN. The detailed quantitative results are

summarized in Table 6-9, and Figure 6-10 with comparative reference to the previous ANN (Table 6-5) and CNN (Table 6-6) models.

The hybrid model achieved a total RMSE of 12.89 kWh for hourly predictions, comparable to the CNN (12.66 kWh) and notably better than the ANN (14.94 kWh). This confirms that incorporating a convolutional head enhanced the model's capability to reproduce fine-scale temporal variations while maintaining strong generalization.

Table 6-9- Evaluation metrics for Hybrid ANN-CNN performance on the unseen test dataset

Output metric	RMSE	MAE	R²
EUI	1.08 kWh/m ²	0.82 kWh/m ²	0.998
TEDI	0.65 kWh/m ²	0.48 kWh/m ²	0.995
CEDI	1.33 kWh/m ²	0.96 kWh/m ²	0.996
Hourly	12.891 kWh	9.033 kWh	0.995

For the annual performance indicators, the hybrid model retained the ANN's high accuracy, with R² values exceeding 0.995 for all three metrics (EUI, TEDI, and CEDI). Among these, the hybrid achieved the lowest RMSE for TEDI (0.65 kWh/m²), surpassing both ANN (0.75 kWh/m²) and CNN (0.99 kWh/m²), confirming that the addition of convolutional layers did not compromise aggregated annual accuracy.

Figure 6-10 presents the relationship between numerical vs surrogate for the hybrid ANN–CNN model evaluated on the unseen test dataset. For EUI, Figure 6-10(a), the predicted points remain closely clustered around the unity line across the full range of values, indicating strong agreement between model predictions and simulation outputs and demonstrating the model's robustness in capturing overall annual energy performance.

Figure 6-10(b) and (c) for TEDI and CEDI, respectively, has somewhat dens distribution close to the unity line. However, dispersion of surrogate values increases at higher index magnitudes, with points deviating further from the perfect regression line. This trend suggests larger prediction errors for heating- or cooling-dominated cases. It should be noted that these higher TEDI and CEDI values correspond to building configurations that fall outside the performance thresholds typically associated with high-performance or zero-carbon-ready buildings under CAGBC guidelines. As such, the observed deviations occur primarily in regions that are less critical for the intended application domain of this surrogate model.

Figure 6-10(d) compares the probability density distributions of the predicted and true hourly energy consumption values for the hybrid ANN–CNN model on the test dataset. The close overlap between the predicted and reference distributions indicates that the model successfully captures the statistical characteristics of hourly energy use, including central tendency, spread, and overall distribution shape.

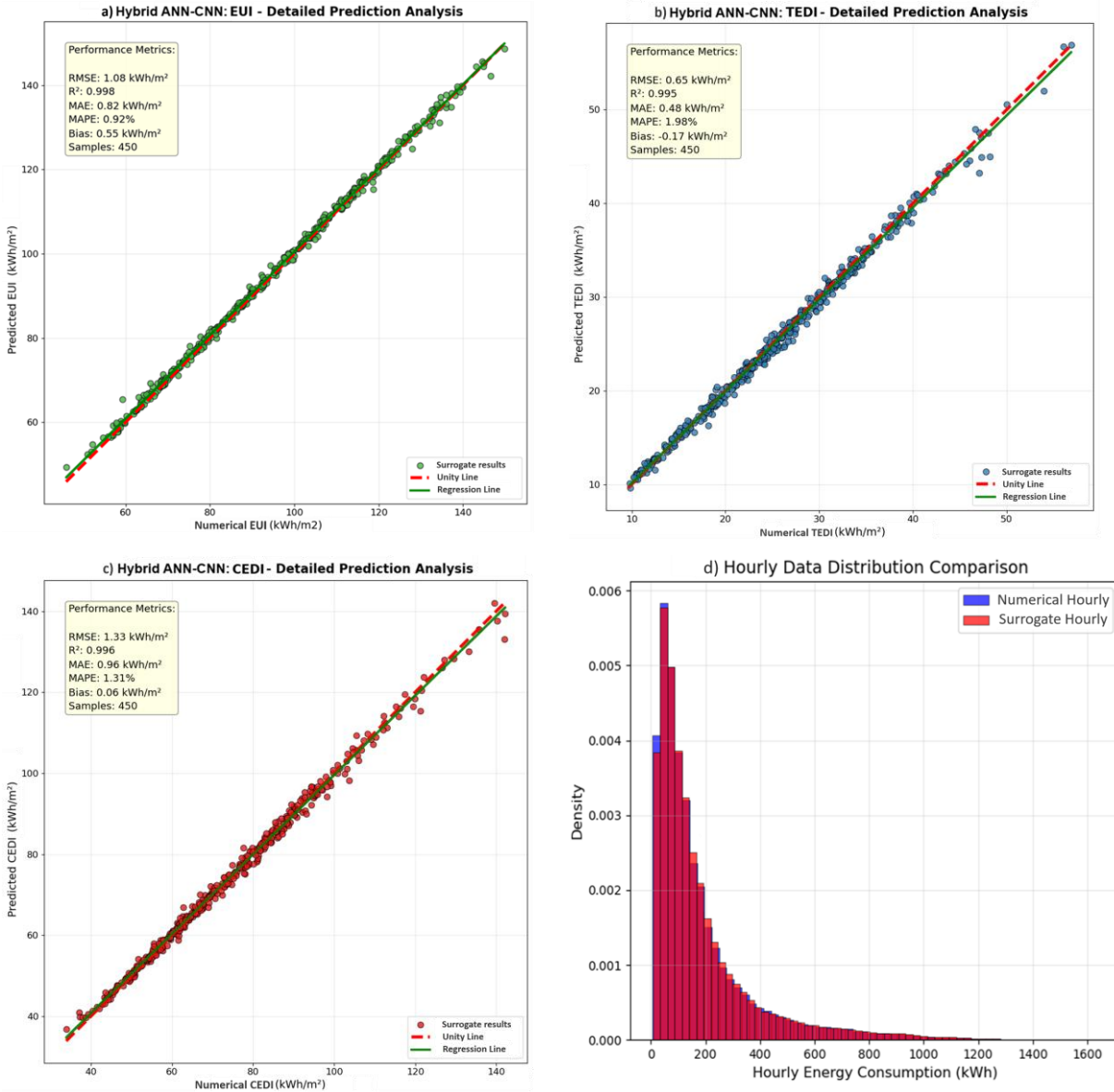


Figure 6-10- Numerical versus Surrogate for hybrid ANN-CNN model evaluation on test dataset

Unlike point-wise regression plots, this density-based comparison emphasizes the model's ability to reproduce the full range and frequency of hourly energy demand patterns rather than individual time-step accuracy alone. The strong agreement between the two distributions confirms that the convolutional

output head effectively learns the temporal structure and variability of hourly consumption, supporting the model's suitability for downstream applications such as clustering, temporal pattern analysis, and operational behavior interpretation. Importantly, the hybrid model achieved this level of accuracy with a moderate computational demand, significantly lower than the CNN's full convolutional training time and only slightly above that of the simple ANN. This efficiency gain is attributed to the hybrid model's design, where the ANN layers capture global nonlinear relationships among scalar input parameters, while the convolutional head processes the structured hourly outputs, effectively learning temporal coherence without incurring the cost of deep convolutional stacks.

In summary, the **hybrid ANN–CNN model** achieved the most balanced performance profile, among the evaluated models, by matching or exceeding the ANN's accuracy on annual indicators, approaching the CNN's accuracy on hourly load prediction, and maintaining a favorable computational cost-to-accuracy ratio.

These results suggest that the hybrid architecture as the most effective surrogate modelling approach developed in this study, providing reliable and computationally efficient predictions for both aggregated (annual) and time-series (hourly) energy performance of high-performance office buildings.

6.5 Final Standalone Model Saving and Post-Processing

6.5.1 Standalone Surrogate Model Deployment

Following the optimization, training, and evaluation stages, the finalized hybrid ANN–CNN ensemble model was exported as a standalone Python application to facilitate independent use without re-training. The deployment workflow involved saving both the model architecture and learned weights in serialized formats to ensure reproducibility and transferability.

The standalone deployment includes three primary artifacts. First, the model configuration, comprising the network structure as well as the normalization ranges and scaling factors used for pre- and post-processing, was stored in a JSON file. Second, the trained model parameters were saved in PyTorch format as a .pth file. Third, an accompanying Python script was generated to automatically load the configuration and weights, collect user inputs, and execute inference.

From an end-user perspective, the script provides a command line interface in which an operator enters the required input parameters based on embedded descriptions of each feature, including its name, units, and valid range. After execution, the tool returns two categories of outputs: i) predicted annual performance indicators (EUI, TEDI, and CEDI), printed directly to the command line, and ii) a full-year

hourly energy consumption profile (365×24), exported automatically as a CSV file in the working directory.

To enhance interpretability and high-performance target check, the script further integrates Canadian Green Building Council (CaGBC) Zero Carbon Building standards as a built-in benchmark. After generating the predictions, the model automatically compares the predicted EUI and TEDI values against the required thresholds and reports whether the building meets or fails the standard criteria. This deployment approach aims to transform the trained surrogate model into a ready-to-use analytical tool, capable of supporting both rapid design scenario testing and energy standard compliance screening for design teams.

6.5.2 Post-Processing and Clustering Analysis of Hourly Predictions

While hourly predictions provide detailed time-resolved information, interpreting large volumes of predicted profiles across multiple design iterations remains a practical challenge. To address this, clustering-based representations are incorporated as an auxiliary output of the surrogate framework, enabling compact and interpretable summaries of building energy behavior that support rapid design comparison without detailed post-processing of raw hourly outputs

The case study analysis presented in the Chapter 4 revealed that examining the hourly energy consumption data of the high-performance office provided rich insights into the operational behavior of the building, highlighting daily and seasonal patterns, as well as deviations related to occupancy schedules, mechanical system performance, and external climatic variations. Those findings demonstrated that annual energy metrics alone could not fully capture the building's dynamic energy use characteristics.

Motivated by those observations, the final step in the surrogate modelling workflow involved performing a detailed clustering analysis on the hourly energy predictions generated by the hybrid ANN–CNN ensemble model. The objective was to replicate the exploratory analytics previously conducted on the office case study dataset using surrogate-predicted outcomes, in order to evaluate whether the surrogate can reproduce meaningful behavioral groupings and capture the underlying temporal trends.

The surrogate model, capable of producing full-year hourly energy profiles for any given input configuration, provided an ideal platform for this post-processing step. Each hourly output dataset (saved automatically as a .csv file when the standalone model is executed) was subjected to unsupervised clustering to categorize the building's operational patterns. The clustering serves as a direct continuation of the case study's analytical framework, transforming predicted data into interpretable operational modes.

Three **K-means clustering configurations**, each corresponding to a distinct k value identified through different selection methods, were applied to the predicted hourly profiles, maintaining consistency with the earlier case study investigation methodology:

1. **Elbow-Determined k (Data-Driven Approach)**

The first configuration adopted the k value determined through the elbow method, which minimizes the within-cluster sum of squares (WCSS). This method provided the most coherent and interpretable clusters in the case study building analysis and was therefore prioritized for the surrogate-predicted data.

2. **Calinski–Harabasz Criterion (Validation Approach)**

To reinforce the robustness of the clustering process, the Calinski–Harabasz (CH) index was also used to estimate k . The CH-based results were compared with those from the elbow method, and in several cases, both identified similar cluster counts, confirming consistency in the identified energy behavior patterns.

3. **Reference Case ($k = 5$) based on the Office Case Study Building Investigation**

In alignment with the 2022 operational year, which represented a stable baseline for the case study building (no pandemic-related disruptions, which was the case in 2021, and no HVAC faults, as observed in 2023) the clustering was repeated with $k = 5$. This configuration facilitated direct comparison between the surrogate-generated patterns and the actual measured data under normal operation conditions.

Although the Silhouette index was also examined, its first maximum typically occurred at $k = 2$, which proved too coarse to differentiate between distinct energy behavior modes. Therefore, the elbow and CH methods were prioritized for selecting representative k values.

It should be noted that, depending on the dataset characteristics, the different k selection methods may yield identical or very similar values for k , resulting in same clustering patterns appearing under different configuration names in the outputs.

In addition to the conventional cluster scatter plots and average daily load profiles, a calendar-map visualization was generated for each k configuration. These calendar maps color-code the dominant cluster for each day of the year, offering a clear temporal representation of how operational modes shift over days and seasons. This visualization technique initially used under the case study investigation, proved equally effective when applied to the model-predicted hourly data, enabling a direct qualitative comparison between measured and surrogate-generated patterns.

6.5.2.1 Illustrative Example: Clustering Analysis for a Random Office Building

For visualizing the k-means clustering process on the hourly output data, one randomly selected input configuration (showed in Table 6-10) was provided to the stand-alone hybrid ANN–CNN model, and the resulting clustering outputs are presented in Figure 6-11.

Table 6-10- A set of random input parameters to perform clustering analysis

Parameter name for the sample (Unit)	Input Value
Building Footprint Area (m ²)	962.205
Floor Plate Aspect Ratio; L/W (-)	0.507
Window Thermal Conductance; U-factor (W/m ² ·K)	0.612
Window Solar Heat Gain Coefficient; SHGC (-)	0.680
Building Orientation Relative to True North (°)	152.42
Wall Thermal Resistance (m ² ·K/W)	5.41
Roof Thermal Resistance (m ² ·K/W)	10.47
Infiltration per Exterior Wall Area (m ³ /s·m ²)	1.72×10^{-4}
Window-to-Wall-Ratio; WWR (-)	0.197
Lighting Power Density; LPD (W/m ²)	7.462
Equipment Power Density; EPD (W/m ²)	9.077
Number of middle floors; N_of_Fl (floors, integer)	2

The first observation shown in Figure 6-11 is that for $k = 4$ (Elbow method), all weekend days are grouped into a single cluster (magenta). However, when the number of clusters increases to $k = 5$ (fixed-k configuration), weekend days are divided into two distinct clusters (magenta and purple). The higher-consumption weekend cluster (magenta) occurs predominantly during the summer months. This pattern indicates that the building requires higher energy input during hot summer weekends to offset internal equipment loads and cooling demand.

For this input configuration, the annual indices presented in Table 6-11, together with the representative daily load profiles (Figure 6-11), clearly indicate that the building is cooling-dominated. Considering weekday behavior, the highest energy demand occurs during the summer months (cyan cluster), while the lowest weekday consumption is observed in winter (red cluster). The seasonal separation of weekday energy behavior is therefore well reflected in the clustering structure.

When k is further increased to 6 (Calinski–Harabasz method), the weekend grouping remains consistent with the $k = 5$ configuration, and the additional cluster is primarily formed from weekday patterns. However, this added fragmentation produces a more complex calendar map and provides limited

additional insight into the building's energy behavior, suggesting diminishing interpretability beyond five clusters.

Table 6-11- Annual energy indices for the random office building

EUI	TEDI	CEDI
96.75	19.24	81.18

Overall, this example demonstrates that applying multiple clustering configurations allows for a more comprehensive understanding of the building's operational patterns, with each k value providing complementary perspectives on the underlying energy behavior.

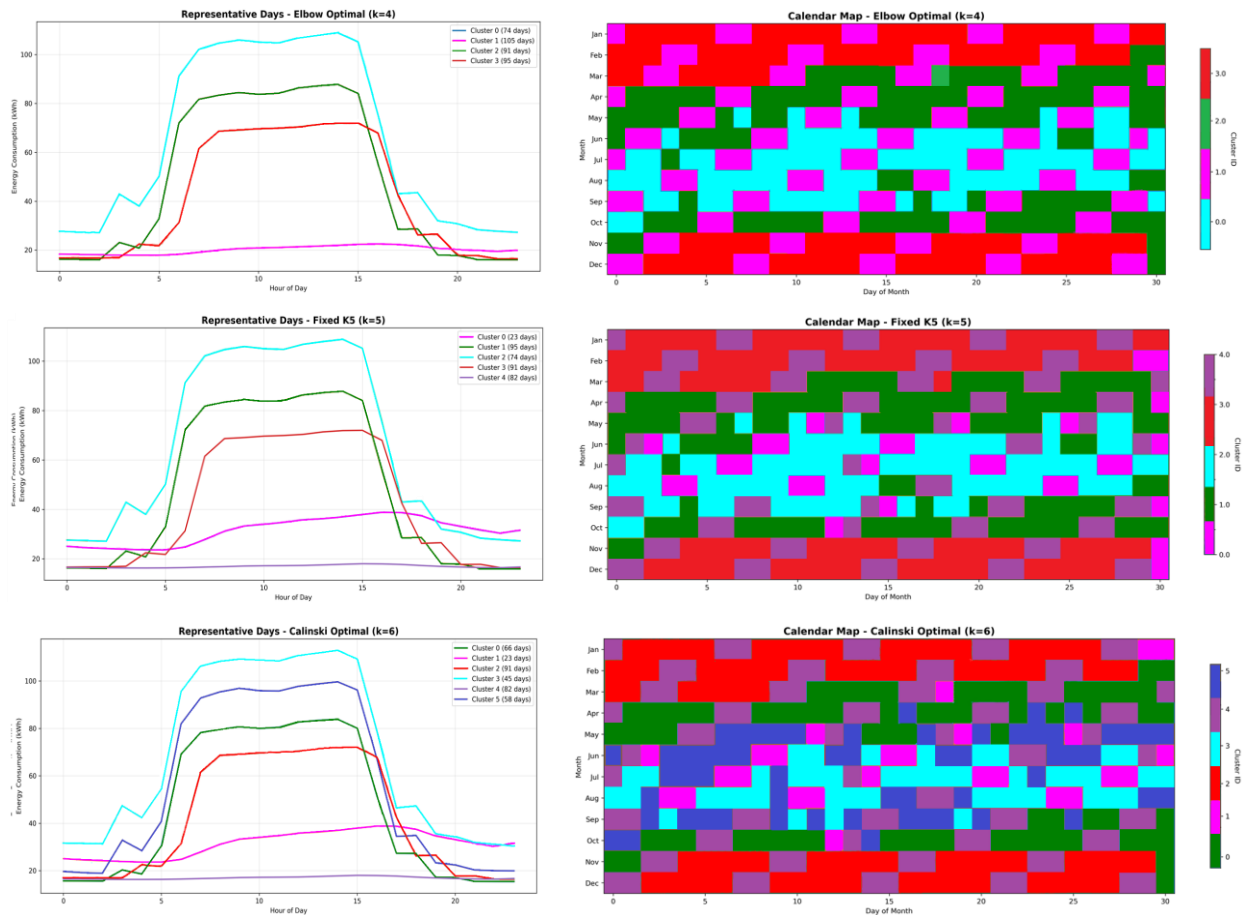


Figure 6-11- Representative daily load profiles and calendar maps from K-means clustering of the hybrid ANN–CNN model's hourly predictions for a randomly selected input configuration, showing results for the elbow, Calinski–Harabasz, and fixed $k = 5$ configurations.

6.5.2.2 Clustering Analysis for an Input Configuration that Represents the Case Study Office

Following the random office case, a clustering analysis was performed for a surrogate model input configuration intended to represent the office case study building from Chapter 4. The selected input

parameters for this configuration are summarized in Table 6-12 while the clustering results (i.e., daily profile representatives and calendar maps) for the three different k values are presented in Figure 6-12. The input parameters for this configuration were selected based on the methodology described in Chapter-5 for Sensitivity Analysis for the Office Case Study.

Table 6-12- Surrogate input values to the case study building

Parameter name for the sample (Unit)	Input Value
Building Footprint Area (m ²)	3135
Floor Plate Aspect Ratio; L/W (-)	0.344
Window Thermal Conductance; U-factor (W/m ² ·K)	0.87
Window Solar Heat Gain Coefficient; SHGC (-)	0.42
Building Orientation Relative to True North (°)	120
Wall Thermal Resistance (m ² ·K/W)	5.28
Roof Thermal Resistance (m ² ·K/W)	7.04
Infiltration per Exterior Wall Area (m ³ /s·m ²)	2.55×10^{-4}
Window-to-Wall-Ratio; WWR (-)	0.37
Lighting Power Density; LPD (W/m ²)	3.72
Equipment Power Density; EPD (W/m ²)	5.51
Number of middle floors; N of Fl (floors, integer)	1

As shown in Figure 6-12, and consistent with the case study analysis, the clustering results of the surrogate-generated data reveal stable weekend behavior across different values of k . Regardless of the selected number of clusters, a single weekend cluster with relatively constant energy consumption throughout the year (magenta cluster) is identified. This indicates that the surrogate model successfully captures the reduced operational variability during weekends.

The maximum energy consumption in the surrogate dataset occurs during winter (red cluster), reflecting heating-dominated peak demand under this input configuration. Summer consumption is represented by the cyan cluster in the $k = 3$ and $k = 4$ solutions. When the number of clusters increases to $k = 5$, the summer cluster splits into two separate groups; however, this subdivision does not reveal a clearly interpretable seasonal pattern in the calendar map.

One cluster observed in the case study results, the black cluster representing elevated energy demand on the first weekday following winter weekends, does not appear in the surrogate model results. In the measured data, this behavior is attributed to increased heating demand required to recondition the building's thermal mass after prolonged weekend setback periods. However, in the numerical model developed for data generation, most envelope constructions are modeled using no-mass material layers, resulting in limited thermal storage capacity. Consequently, the model does not reproduce this thermal mass-driven recovery effect. This discrepancy highlights a limitation of the current parametric modelling

assumptions and suggests that incorporating thermal mass as an input parameter could further improve surrogate prediction capabilities.

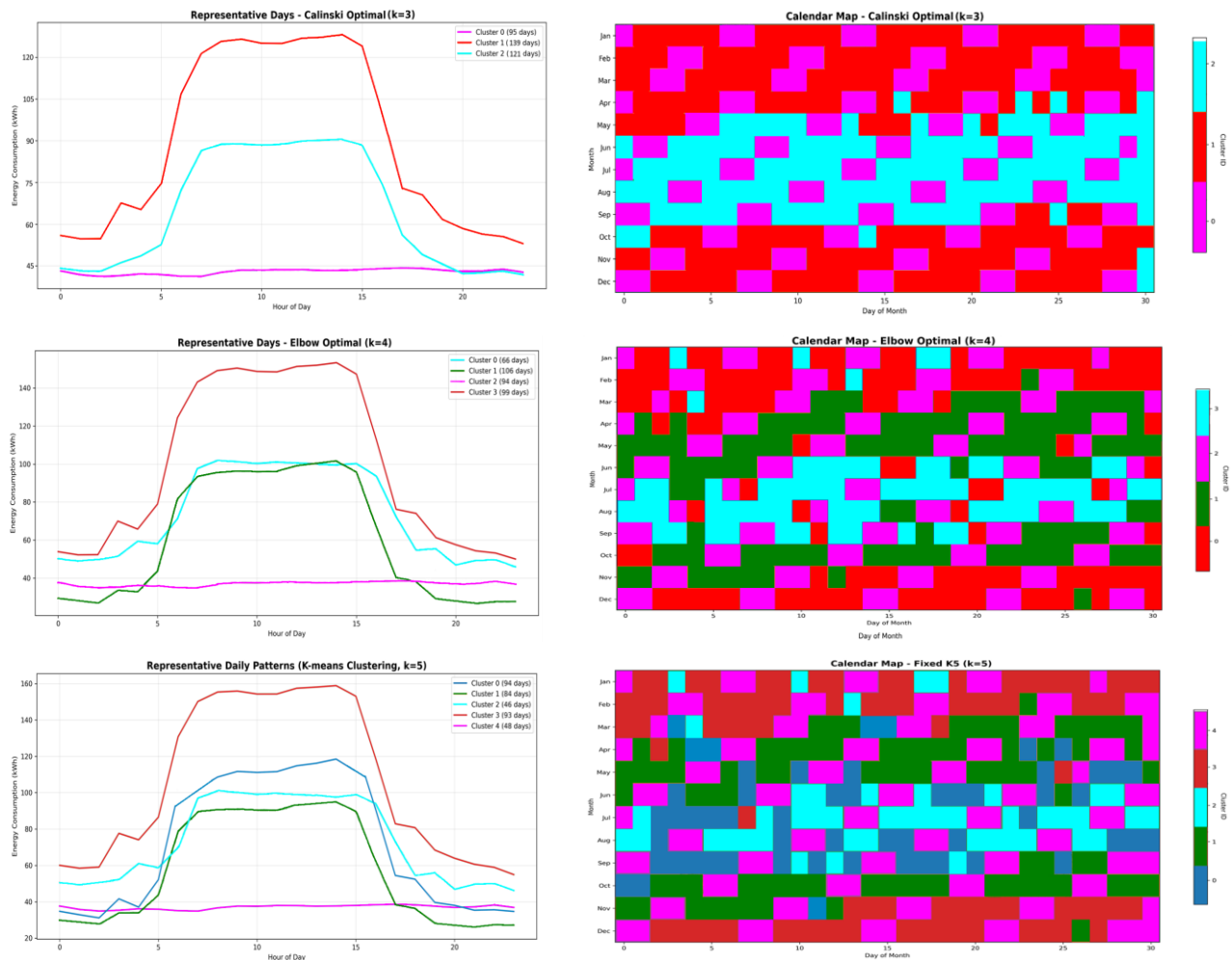


Figure 6-12- Representative daily load profiles and calendar maps from K-means clustering of the hybrid ANN–CNN model’s hourly predictions for a closest input values to the case study building, showing results for the elbow, Calinski–Harabasz, and fixed $k = 5$ configurations.

6.6 Sensitivity Analysis of Surrogate Model

6.6.1 Overview and Methodology

A comprehensive sensitivity analysis was conducted to quantify how building design parameters influence the surrogate model outputs, with primary emphasis on hourly energy consumption predictions. Four complementary statistical and model-based methods were applied across multiple temporal resolutions (hourly, daily, and annual) to support a robust interpretation of parameter effects.

6.6.2 Correlation Analysis (Baseline)

The magnitude of the correlation coefficient $|r|$ indicates the strength of the linear association between an input parameter and the output, with values closer to 1 (or -1) reflecting stronger positive (or negative) relationships, and values near 0 indicating weak linear association. However, correlation does not imply causality and may not capture nonlinear effects or interactions. Therefore, these results are used as a first-order baseline and are complemented by the subsequent sensitivity analyses. Figure 6-13 shows the correlation coefficients for both the annual energy performance metrics and the total daily energy consumption. The analysis indicates that building size parameters, specifically, the building floor area and the number of middle floors, are the dominant factors influencing short-term (daily) energy consumption patterns in the surrogate model.

When it comes to annual performance metrics (EUI, TEDI, and CEDI), the EPD emerged as the most influential parameter across all outputs, except for TEDI. In the case of TEDI, infiltration rate showed the strongest correlation, with EPD being the second most impactful variable. These findings highlight that while spatial and geometric factors primarily drive short-term load variability, equipment-related and envelope performance parameters exert greater influence on aggregated annual energy outcomes.

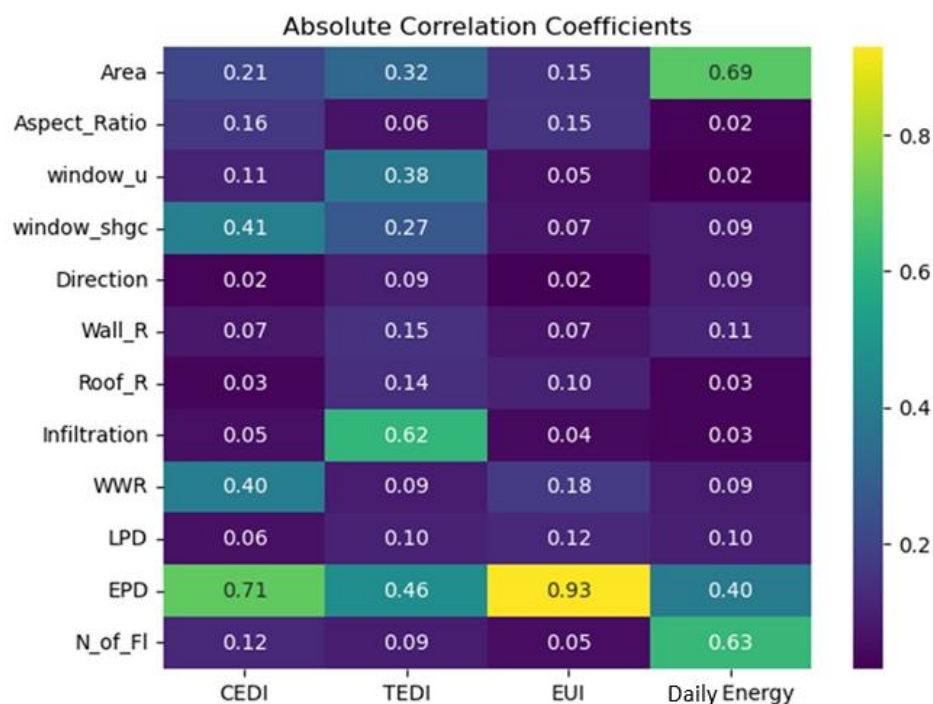


Figure 6-13- Correlation coefficients for annual energy metrics and total daily energy

6.6.3 One-at-a-Time (OAT) Analysis

The One-at-a-Time (OAT) sensitivity analysis was conducted to quantify the individual influence and directional effect of each input parameter on both annual and hourly model outputs. In this approach, each

input variable was perturbed independently (e.g., $\pm 10\%$ change from its baseline value) while all other inputs were held constant. The resulting relative percentage change in the model output was then computed to obtain the OAT sensitivity score.

The magnitude of the OAT score represents the relative change in output induced by perturbing a single parameter. Larger absolute percentage changes indicate stronger local sensitivity, while the sign of the change reflects the direction of influence (positive values indicate that increasing the parameter increases the output, whereas negative values indicate an inverse relationship). Because this method evaluates parameter influence around a reference configuration, it characterizes local behavior only and does not account for parameter interactions or nonlinear effects across the full design space. Therefore, OAT results are interpreted as indicators of local influence and are complemented by global and model-based sensitivity analyses in subsequent sections.

Figure 6-14 presents the OAT results for annual energy performance indicators and the hourly energy outputs, respectively. For the annual metrics, the results are expressed as signed percentage effects, illustrating whether an increase in a given parameter leads to a positive (increase) or negative (decrease) change in energy use. For instance, higher envelope insulation or reduced infiltration rates exhibit reduction effects on annual EUI, indicating improved energy performance, whereas increases in equipment (EPD) or lighting power (LPD) densities produce positive effects, reflecting higher overall consumption.

In contrast, the OAT results for the hourly energy data are presented as normalized relative magnitudes rather than signed effects. This is because hourly energy responses exhibit complex temporal variations, parameter changes can increase energy use during certain periods (e.g., cooling season) and decrease it during others (e.g., heating season). As such, the hourly OAT plot reflects the overall strength of each parameter's influence on short-term energy dynamics, without assigning a directional sign.

Based on the OAT sensitivity analysis, building floor area (Area), equipment power density (EPD), and number of middle floors (N_of_Fl) were identified as the most influential input parameters affecting the hourly energy outputs, reflecting their strong impact on short-term load variability. Among the annual performance indicators, EPD emerged as the most dominant parameter across all three metrics (EUI, TEDI, and CEDI), emphasizing its consistent influence on overall building energy performance.

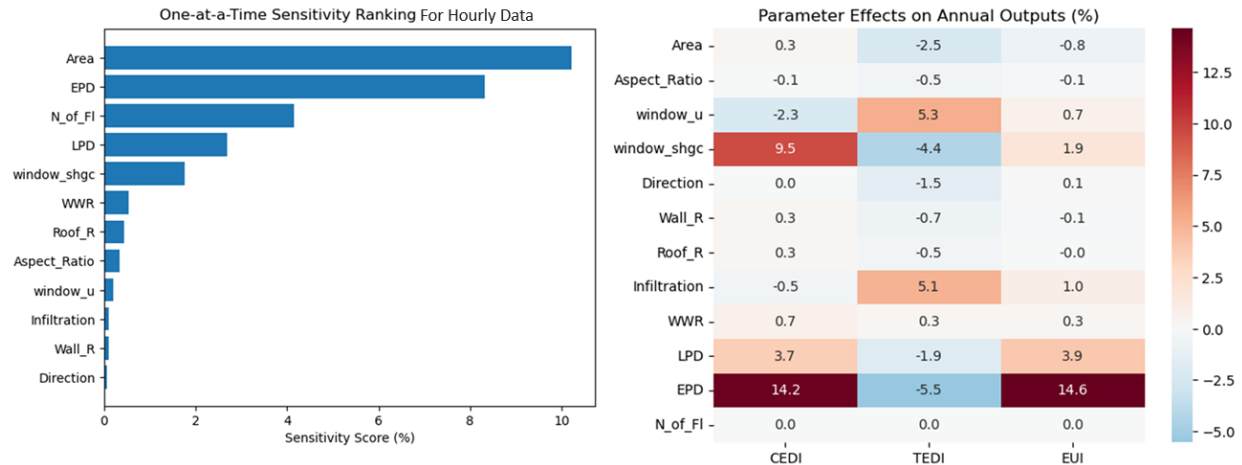


Figure 6-14- OAT results for hourly energy outputs and annual energy performance indicators

6.6.4 Morris Elementary Effects (Global Sensitivity Analysis)

To further capture nonlinear dependencies and parameter interaction effects beyond those detectable by the OAT approach, the Morris Elementary Effects (EE) method was applied. This global sensitivity method evaluates how each input parameter influences the model's hourly energy output predictions across the entire input domain. Unlike local methods such as OAT or correlation analysis, the Morris approach perturbs each input along multiple trajectories within the parameter space, enabling assessment of both overall importance and interaction behavior.

The mean absolute elementary effect (μ^*) represents the overall influence of a parameter on the model output, with larger μ^* values indicating stronger global sensitivity. The standard deviation of the elementary effects (σ) reflects the degree of nonlinearity and interaction with other variables; higher σ values suggest stronger interaction effects or nonlinear behavior, whereas lower σ values indicate more linear and independent influence. Parameters characterized by high μ^* and high σ are therefore considered both influential and interactive, while low values for both metrics suggest limited impact on model behavior.

Although the Morris method provides a broader global perspective compared to local sensitivity approaches, it remains a screening method and does not quantify exact variance contributions. Therefore, its results are interpreted as indicators of global importance and interaction tendencies, and are complemented by permutation-based and composite sensitivity analyses for comprehensive evaluation.

Figure 6-15 presents the *Morris Importance vs. Nonlinearity (Interaction)* plot, where each point corresponds to an individual input parameter. Parameters positioned toward the top-right quadrant, such as *Area* and *N_of_Fl*, are both highly influential and strongly interactive, meaning their impacts depend on the combined behavior of other parameters. Inputs near the bottom-left quadrant have limited

influence and behave almost linearly, while parameters aligned closer to the x-axis but low on the y-axis are important yet predominantly linear, showing stable and predictable effects on model performance. This analysis highlights three dominant parameters governing hourly energy variations: Building Floor Area, Number of middle Floors (N_of_Fl), and Equipment Power Density (EPD).

Specifically, Building Floor Area shows the highest importance and strong interaction effects, positioned at the top-right corner of the plot. This reflects its compounded influence on overall thermal mass, conditioned volume, and system energy response. Number of middle Floors (N_of_Fl) also displays high sensitivity and pronounced nonlinearity, indicating that changes in building height alter vertical load distribution and airflow dynamics in a nonlinear fashion. EPD maintains significant importance with moderate nonlinearity, suggesting that equipment-related internal gains interact with other parameters (such as lighting or infiltration) to modulate hourly cooling and plug-load fluctuations.

Meanwhile, parameters such as Lighting Power Density (LPD), Window-to-Wall Ratio (WWR), Infiltration Rate, and Window Properties (U-value, SHGC) cluster toward the lower-left portion of the plot, showing lower importance and weaker nonlinearity, indicating more direct, approximately linear effects on hourly consumption.

In summary, the Morris analysis demonstrates that geometric (Area, N_of_Fl) and internal load (EPD) parameters dominate short-term building energy behavior, while envelope-related features exert secondary and mostly linear influences. These findings confirm that the hybrid ANN–CNN model captures both primary and interaction-driven sensitivities governing hourly energy dynamics.

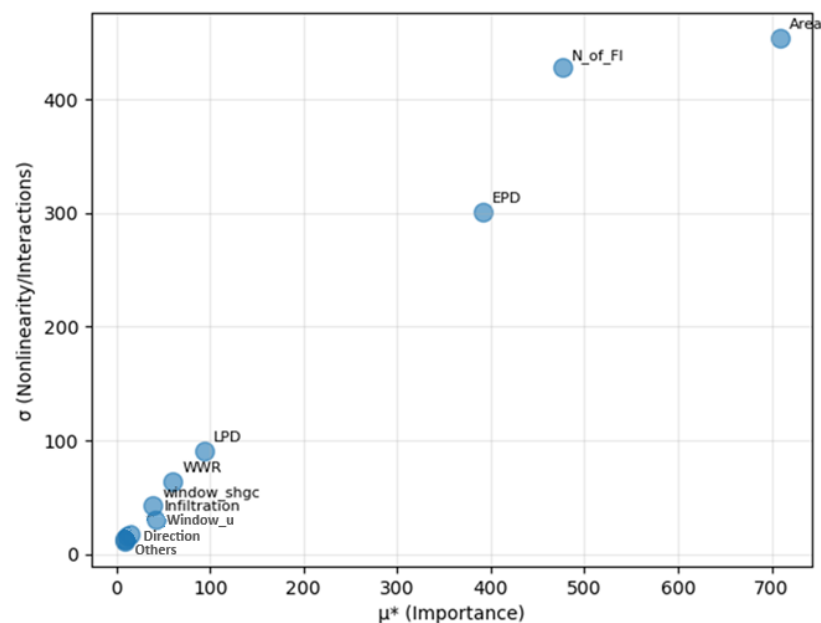


Figure 6-15- Morris sensitivity analysis results for hourly energy outputs

6.6.5 Permutation Importance Analysis

To complement the previous sensitivity analyses and directly assess the influence of each input parameter on the predictive accuracy of the hybrid ANN–CNN surrogate model, the Permutation Importance method was applied. Unlike purely statistical or sampling-based methods, permutation importance evaluates each feature’s contribution to model performance by randomly shuffling its values and measuring the corresponding increase in prediction error, quantified here as the rise in RMSE.

The magnitude of the RMSE increase indicates the relative importance of the parameter: larger increases imply that the surrogate model relies strongly on that input for accurate prediction, whereas minimal change suggests limited predictive contribution. Because this method evaluates sensitivity through degradation of model performance, it reflects the model’s dependency on each feature rather than direct physical causality. In other words, a highly ranked parameter is important for predictive accuracy within the learned surrogate structure, though it may not necessarily represent the strongest physical driver in isolation. Therefore, permutation importance is interpreted as a model-based sensitivity measure, capturing predictive relevance within the trained hybrid architecture, and is considered alongside physics-based and global sensitivity methods to provide a balanced and comprehensive assessment of parameter influence.

Figure 6-16 illustrates the permutation importance ranking for the model’s hourly output predictions. The results show that Equipment Power Density (EPD) exerts the strongest overall impact on model accuracy; permuting this variable caused the largest deterioration in performance, confirming its dominant role in defining energy consumption intensity. The next most influential feature is Building Area which strongly influence the building’s total load magnitude and distribution of solar gains.

A second group of moderately influential inputs includes the number of middle floors ($N_{\text{of_FI}}$), primarily a geometric parameter, along with window SHGC, infiltration rate, lighting power density (LPD), window thermal transmittance, and WWR. These parameters affect envelope thermal performance by balancing solar gains, heat losses, and internal loads. Finally, roof and wall insulation levels, orientation, and aspect ratio exhibit minimal impact on permutations, suggesting that within the training data range, their variations were either captured by correlated features or exerted relatively weak independent effects on hourly energy predictions.

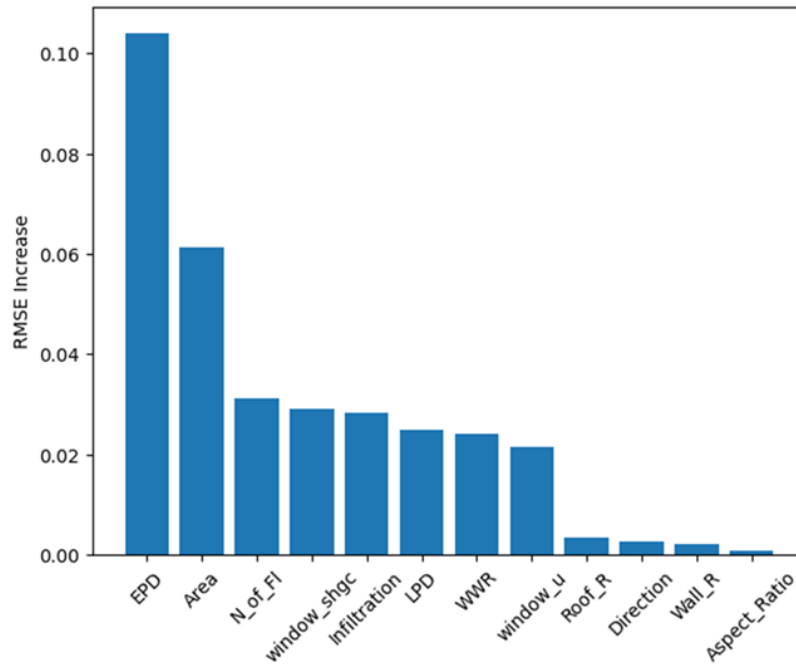


Figure 6-16- Permutation importance ranking for the model's hourly output predictions

Collectively, the permutation importance analysis validates the findings from the previous OAT and Morris analyses: while geometry-related parameters (Area, N_of_Fl) govern overall energy magnitude, operational and load-intensity parameters (EPD, LPD, Infiltration) are critical for determining fine-scale energy variability.

6.6.6 Integrated Sensitivity Framework for Hourly Outputs

To synthesize the results from the four sensitivity and interpretability methods, each metric was first normalized to a 0-1 scale and then combined through simple averaging to produce an integrated sensitivity indicator. This aggregated measure is not intended to represent a physical sensitivity, but rather to highlight parameters that consistently rank as influential across multiple analytical perspectives. By emphasizing convergence among methods with different assumptions and scopes, this integrated sensitivity framework provides a basis for parameter ranking and improves confidence in the interpretation of the surrogate model's hourly behavior

Note that the adoption of min-max normalization for the composite scoring provides a balanced comparison by eliminating scale biases introduced by differing magnitudes (e.g., Morris μ^* vs. RMSE). This approach resulted in more interpretable composite scores, highlighting relative importance without distortion from metric-specific ranges. The results of this integrated sensitivity framework for hourly outputs are presented in Table 6-13. The composite analysis shows that:

- Area, EPD, and N_of_Fl are the primary drivers of energy performance and should be prioritized during calibration and optimization.
- LPD, WWR, and Window SHGC serve as secondary tuning variables, especially for refining thermal–daylight trade-offs.
- Envelope and orientation parameters exhibit low individual sensitivity, though they may still interact nonlinearly with key variables.

This integrated sensitivity framework highlights parameters that are consistently influential across the four complementary methods and provides a consensus-based basis for ranking. In doing so, it supports informed parameter screening, model simplification, and prioritization for subsequent optimization and uncertainty analyses, while recognizing that the aggregated score is intended as an interpretive aid rather than a definitive measure of physical sensitivity.

Table 6-13- Sensitivity ranking summary with the normalized value for all methods and normalized composite score.

Parameter	Correlation (normalized)	OAT (normalized)	Morris μ^* (normalized)	Permutation (normalized)	Composite Score
Area	1.000	1.000	1.000	0.574	0.894
EPD	0.550	0.808	0.543	1.000	0.725
N_of_Fl	0.897	0.405	0.669	0.274	0.561
LPD	0.117	0.257	0.121	0.223	0.179
WWR	0.112	0.050	0.073	0.213	0.112
Window SHGC	0.107	0.165	0.044	0.259	0.144
Infiltration	0.017	0.012	0.046	0.254	0.082
Window Thermal Transmittance	0.000	0.024	0.008	0.186	0.055
Wall Thermal Resistance	0.138	0.037	0.007	0.023	0.051
Aspect Ratio	0.100	0.024	0.011	0.019	0.039
Direction	0.104	0.012	0.005	0.000	0.030
Roof Thermal Resistance	0.017	0.037	0.000	0.008	0.016

Chapter 7

7 Summary and Conclusions

This chapter synthesizes the findings presented across Chapters 4, 5, and 6, draws conclusions against the research questions stated in Chapter 1, and identifies directions for future research.

7.1 Purpose and Scope of the Study

The primary objective of this research was to develop and validate a computationally efficient modelling framework capable of accurately predicting both annual energy performance indicators and high-resolution hourly energy consumption profiles for institutional high-performance buildings. While surrogate modelling has been widely applied in building energy research, most existing studies emphasize scalar outputs such as annual Energy Use Intensity (EUI), with limited attention to structured temporal prediction and physical interpretability.

To address this gap, the study developed a surrogate modeling framework for predicting the energy behavior of high-performance office buildings, integrating insights from the case study analysis and using machine learning techniques trained on simulation-generated data. The research integrates three methodological components that support this objective:

1. **Case study analysis** of a high-performance office building using clustering–regression techniques to identify dominant operational energy patterns and establish the empirical basis for the modelling assumptions adopted in the parametric framework,
2. **Development of a hybrid deep learning surrogate model (ANN–CNN)** capable of mapping building design and operational parameters to high-resolution energy outputs, and
3. **Sensitivity analysis applied to the surrogate model**, to evaluate parameter influence and confirm that the learned relationships remain physically plausible.

The scope of the work also included the development of a parametric EnergyPlus simulation framework for dataset generation, structured sampling of the design space, surrogate model training and validation, clustering-based evaluation of temporal energy patterns, and sensitivity assessment conducted directly on the trained surrogate model.

7.2 Summary of Key Findings

7.2.1 Case Study and Clustering–Regression Analysis

The office case study investigation demonstrated that the clustering–regression framework provides a simple, fast, and accurate representation of daily load behavior in the high-performance net-positive institutional case study building. The approach successfully identified recurring seasonal and operational energy patterns and enabled rapid prediction suitable for integration into building control systems. This confirms that clustering provides a meaningful abstraction of complex hourly data, enabling both simplified interpretation and identification of operational regimes that would not be evident from raw time-series analysis.

Clustering analysis of the measured building data revealed the emergence of previously unseen daily energy consumption patterns during later operational periods, suggesting that the method may also support early identification of abnormal operating conditions or potential faults. Analysis of the clustered load profiles indicated that building and operational parameters affecting individual load components, particularly HVAC operation, plug loads, and lighting usage patterns, play a larger role in shaping the daily energy profile than short-term variations in outdoor weather conditions, indicating that internal operational drivers dominate energy behavior in high-performance buildings. In the parametric modelling framework, these operational effects are represented through parameters such as EPD and LPD. These findings highlight the importance of internal loads and system operation in determining the energy behavior of high-performance office buildings, where improved envelope performance reduces the relative influence of external climatic conditions.

7.2.2 Surrogate Modelling Framework

The hybrid ANN–CNN surrogate architecture achieved very high predictive accuracy, with a minimum R^2 of 0.995 across evaluated outputs. The model reproduced both annual performance indices (EUI, TEDI, CEDI) and full 365×24 hourly energy profiles across a broad parametric design space.

Unlike many surrogate studies that focus primarily on aggregated annual metrics, this framework demonstrated that high-resolution temporal outputs can be learned with strong fidelity when model architecture aligns with output structure. The comparative analysis of ANN, CNN, and hybrid architecture further demonstrates that model performance is strongly dependent on alignment between model structure and output representation. While ANN models are effective for aggregated outputs, CNN-based architectures improve the prediction of structured hourly profiles by capturing temporal dependencies. The hybrid model combines these strengths, achieving balanced performance across both output scales.

Clustering analysis of surrogate-generated profiles confirmed preservation of seasonal transitions and daily load shapes, indicating that the model maintains physically meaningful temporal organization rather than merely minimizing statistical errors.

7.2.3 Sensitivity Analysis and Physical Verification

Sensitivity analysis was performed directly on the trained surrogate model to investigate how variations in input parameters influence predicted building energy performance. Across the different analysis approaches, including correlation analysis, One-at-a-Time (OAT) perturbations, Morris elementary effects, and permutation importance, consistent parameter rankings were observed. Internal loads and geometric parameters were identified as dominant drivers of performance variation, while envelope thermal improvements exhibited diminishing marginal influence under cooling-dominated conditions.

To further assess the reliability of the surrogate predictions, the sensitivity results were examined against established building-physics expectations. The observed relationships between parameters and energy outcomes were consistent with known physical behavior, providing qualitative verification that the surrogate model captured plausible system responses rather than purely statistical correlations. The convergence of results across multiple sensitivity methods strengthened confidence in the surrogate model's ability to represent the relative influence of key design and operational parameters.

Taken together, these findings answer the two research questions posed in Chapter 1. RQ1 is addressed through the clustering-regression analysis, which identifies EPD, LPD, and HVAC operation as dominant drivers of time-resolved energy behavior in high-performance office buildings. RQ2 is addressed through the hybrid ANN-CNN framework, which demonstrates that hourly energy profiles can be predicted with $R^2 \geq 0.995$ from a 12-parameter design input space, confirming that structured surrogate modelling of high-resolution temporal outputs is both feasible and accurate.

7.3 Contributions

The contributions of this research are situated within the domain of building energy modelling and performance prediction, rather than the development of novel machine learning methodologies. The study applies established machine learning techniques in a structured and domain-informed manner to address challenges specific to high-performance buildings, particularly the prediction of high-resolution temporal energy behavior. As such, the value of this work should be interpreted in the context of its contribution to energy modelling practice, including improved interpretability, integration with physical insights, and applicability to design and operational decision-making. Within this context, this thesis makes the following primary contributions to building energy modelling practice:

1. Structured Surrogate Modelling of High-Dimensional Temporal Outputs

A hybrid ANN-CNN framework was developed to map low-dimensional design parameters to high-dimensional hourly energy sequences. This addresses a limitation frequently observed in prior surrogate modelling studies that concentrate primarily on annual or monthly metrics.

2. Integration of Simulation, Machine Learning, and Physical Verification

A structured and reproducible workflow was developed that combines parametric simulation, systematic sampling, surrogate model training, and hypothesis-based physical consistency checks. This framework provides a transparent linkage between physics-based modelling and data-driven prediction.

3. Temporal Structure Validation Through Clustering Analysis

The study demonstrated that surrogate-generated outputs preserve interpretable seasonal and operational load structures, providing structural validation beyond traditional error metrics.

Collectively, these contributions directly address the gaps identified in the literature review of Chapter 2: the limited focus on high-performance buildings, the predominance of low-resolution surrogate outputs, the restricted treatment of design parameters in hourly models, and the absence of structured physics-to-surrogate pipelines.

7.4 Implications for Design and Operation

The proposed methodological framework and its findings have direct implications for both building design and operational management.

Designers can rapidly assess geometric configurations, internal load strategies, and envelope variations without repeated high-cost simulations, while accessing both detailed hourly energy profiles and clustering-based representations of these profiles. The clustered outputs provide an interpretable overview of operational patterns, while the full hourly data preserves critical information such as peak heating and cooling demands, enabling more comprehensive performance assessment.

From an operational standpoint, the clustering–regression framework further supports identification of recurring load structures and detection of deviations from expected behavior. Its ability to reproduce full annual hourly profiles makes it suitable for demand response evaluation, peak load analysis, and grid-interactive building strategies, while the clustering representation facilitates rapid interpretation of system performance.

More broadly, the research demonstrates that data-driven models can complement physics-based simulation in a manner that preserves interpretability and physical coherence, supporting performance-informed decision-making.

7.5 Limitations

Several limitations define the applicability domain of this study. These limitations primarily reflect deliberate scope decisions and modelling simplifications adopted to ensure tractability and clarity of

analysis. Importantly, they do not represent fundamental restrictions of the proposed framework, but rather define the current implementation boundaries, which can be systematically extended in future work.

The surrogate model was developed for a single building typology, a high-performance office building in Waterloo, ON, with rectangular geometry and a fixed VRF-based HVAC configuration with ERV-supported ventilation. Alternative geometries, typologies, mechanical systems or geographic locations may exhibit different sensitivity structures and operational behaviors.

Operational schedules were not parameterized as structured inputs. Given that schedule effects were shown to influence performance significantly in the case study, incorporating time-dependent schedule variability would enhance realism but increase input dimensionality and modelling complexity.

The parametric model employed simplified envelope constructions with limited thermal mass representation, restricting replication of certain transient behaviors such as post-setback thermal recovery.

Sensitivity results remain dependent on defined parameter ranges and climate context. Different climatic conditions or expanded parameter bounds may alter influence rankings.

Finally, the framework focused exclusively on operational energy metrics and did not incorporate cost, carbon, comfort, or uncertainty quantification dimensions. Extending the framework to incorporate these dimensions represents a natural progression for future work.

7.6 Future Work

Based on the aforementioned limitations, future research can extend this framework in several directions. First, incorporating structured schedule variability as vector-based or categorical inputs would improve operational realism. Second, expanding the dataset to include additional building typologies, geometries, and alternative HVAC typologies, including geothermal systems, would enhance generalizability. Third, extending the surrogate toward multi-objective prediction, including life-cycle cost, carbon emissions, and occupant comfort metrics, would broaden decision-support capability. Fourth, integrating uncertainty quantification methods would enable confidence estimation and risk-aware design evaluation. Finally, transfer learning strategies may allow adaptation of the trained surrogate to additional buildings with limited incremental simulation data.

7.7 Conclusions

This research demonstrates that high-resolution surrogate modelling of institutional high-performance buildings is feasible when physics-based simulation, structured sampling, and appropriately designed deep learning architectures are systematically integrated.

The study closes a practical gap between low-dimensional parametric design inputs and high-dimensional temporal energy outputs by demonstrating that hourly energy profiles can be predicted with high fidelity while preserving physically interpretable structure. Rather than replacing physics-based simulation, the proposed framework complements it by enabling rapid scenario evaluation and structured performance analysis.

The central takeaway of this thesis is that surrogate models, when carefully designed and physically validated, can evolve from simple regression approximators into reliable analytical tools capable of supporting both design exploration and operational insight in high-performance buildings.

Collectively, the work contributes to a scalable and computationally efficient methodology that advances incrementally the integration of machine learning and building energy simulation while maintaining physical consistency, interpretability, and practical relevance.

References

- [1] G. Q. Chen *et al.*, “Low-carbon building assessment and multi-scale input–output analysis,” *Commun. Nonlinear Sci. Numer. Simul.*, vol. 16, no. 1, pp. 583–595, Jan. 2011, doi: 10.1016/j.cnsns.2010.02.026.
- [2] “Paris Agreement,” United Nations (UN), Geneva, Switzerland, 2015. [Online]. Available: https://unfccc.int/files/essential_background/convention/application/pdf/english_paris_agreement.pdf
- [3] “ZERO CARBON BUILDING, PERFORMANCE STANDARD VERSION 2,” Canada Green Building Council, Standard 978-1-7781454-3-8, Jun. 2022.
- [4] D. D’Agostino and D. Parker, “Data on cost-optimal Nearly Zero Energy Buildings (NZEBs) across Europe,” *Data Brief*, vol. 17, pp. 1168–1174, Apr. 2018, doi: 10.1016/j.dib.2018.02.038.
- [5] L. Chen *et al.*, “Strategies to achieve a carbon neutral society: a review,” *Environ. Chem. Lett.*, vol. 20, no. 4, pp. 2277–2310, Aug. 2022, doi: 10.1007/s10311-022-01435-8.
- [6] A. Ali, R. Jayaraman, A. Mayyas, B. Alaifan, and E. Azar, “Machine learning as a surrogate to building performance simulation: Predicting energy consumption under different operational settings,” *Energy Build.*, vol. 286, p. 112940, May 2023, doi: 10.1016/j.enbuild.2023.112940.
- [7] T. Østergård, R. L. Jensen, and S. E. Maagaard, “Building simulations supporting decision making in early design – A review,” *Renew. Sustain. Energy Rev.*, vol. 61, pp. 187–201, Aug. 2016, doi: 10.1016/j.rser.2016.03.045.
- [8] D. B. Crawley *et al.*, “EnergyPlus: creating a new-generation building energy simulation program,” *Energy Build.*, vol. 33, no. 4, pp. 319–331, Apr. 2001, doi: 10.1016/S0378-7788(00)00114-6.
- [9] ASHRAE 90.1, *ANSI/ASHRAE/IES Standard 90.1- Energy Standard for Buildings Except Low-Rise Residential Buildings*, 2022.
- [10] NECB, *National Energy Code of Canada for Buildings (NECB) 2020*, 2020.
- [11] 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.
- [12] E. Thrampoulidis, G. Mavromatidis, A. Lucchi, and K. Orehounig, “A machine learning-based surrogate model to approximate optimal building retrofit solutions,” *Appl. Energy*, vol. 281, p. 116024, Jan. 2021, doi: 10.1016/j.apenergy.2020.116024.
- [13] E. Ohene, A. P. C. Chan, and A. Darko, “Prioritizing barriers and developing mitigation strategies toward net-zero carbon building sector,” *Build. Environ.*, vol. 223, p. 109437, Sep. 2022, doi: 10.1016/j.buildenv.2022.109437.
- [14] V. J. L. Gan, I. M. C. Lo, J. Ma, K. T. Tse, J. C. P. Cheng, and C. M. Chan, “Simulation optimisation towards energy efficient green buildings: Current status and future trends,” *J. Clean. Prod.*, vol. 254, p. 120012, May 2020, doi: 10.1016/j.jclepro.2020.120012.
- [15] “Multi-Timescale Energy Consumption Management in Smart Buildings Using Hybrid Deep Artificial Neural Networks.” Accessed: Apr. 17, 2026. [Online]. Available: <https://www.mdpi.com/2079-9268/14/4/54>
- [16] “Welcome | TRNSYS : Transient System Simulation Tool.” Accessed: Nov. 07, 2023. [Online]. Available: <https://www.trnsys.com/>
- [17] Y. T. Chae, R. Horesh, Y. Hwang, and Y. M. Lee, “Artificial neural network model for forecasting sub-hourly electricity usage in commercial buildings,” *Energy Build.*, vol. 111, pp. 184–194, Jan. 2016, doi: 10.1016/j.enbuild.2015.11.045.

- [18] N. Shirzadi, D. Lau, and M. Stylianou, “Surrogate Modeling for Building Design: Energy and Cost Prediction Compared to Simulation-Based Methods,” *Buildings*, vol. 15, no. 13, p. 2361, Jan. 2025, doi: 10.3390/buildings15132361.
- [19] E. Markarian *et al.*, “Informing building retrofits at low computational costs: a multi-objective optimisation using machine learning surrogates of building performance simulation models,” *J. Build. Perform. Simul.*, vol. 0, no. 0, pp. 1–17, Jul. 2024, doi: 10.1080/19401493.2024.2384487.
- [20] A. K. Athienitis and L. O’Brien, Eds., *Modeling, design, and optimization of net-zero energy buildings*. in Solar heating and cooling. Berlin: Ernst & Sohn, 2015.
- [21] ASHRAE, “ANSI/ASHRAE Standard 228, Standard Method of Evaluating Zero Net Energy and Zero Net Carbon Building Performance | ashrae.org.” Accessed: Aug. 03, 2024. [Online]. Available: <https://www.ashrae.org/technical-resources/bookstore/ansi-ashrae-standard-228-standard-method-of-evaluating-zero-net-energy-and-zero-net-carbon-building-performance>
- [22] A. J. Marszal *et al.*, “Zero Energy Building – A review of definitions and calculation methodologies,” *Energy Build.*, vol. 43, no. 4, pp. 971–979, Apr. 2011, doi: 10.1016/j.enbuild.2010.12.022.
- [23] CGBC, “ZERO CARBON BUILDING DESIGN STANDARD VERSION 3,” Canada Green Building Council, Standard 978-1-7781454-1-4, Jun. 2022.
- [24] D. Coakley, P. Raftery, and M. Keane, “A review of methods to match building energy simulation models to measured data,” *Renew. Sustain. Energy Rev.*, vol. 37, pp. 123–141, Sep. 2014, doi: 10.1016/j.rser.2014.05.007.
- [25] S. Rouchier, “Statistical modelling for building energy performance,” Accreditation to supervise research, Université savoie mont blanc, 2022. Accessed: Apr. 25, 2026. [Online]. Available: <https://univ-smb.hal.science/tel-04845644>
- [26] X. Zhang, C. Rasmussen, D. Saelens, and S. Roels, “Time-dependent solar aperture estimation of a building: Comparing grey-box and white-box approaches,” *Renew. Sustain. Energy Rev.*, vol. 161, p. 112337, Jun. 2022, doi: 10.1016/j.rser.2022.112337.
- [27] Z. (Jerry) Yu, F. Haghighat, and B. C. M. Fung, “Advances and challenges in building engineering and data mining applications for energy-efficient communities,” *Sustain. Cities Soc.*, vol. 25, pp. 33–38, Aug. 2016, doi: 10.1016/j.scs.2015.12.001.
- [28] P. Howes and N. Crook, “Using input parameter influences to support the decisions of feedforward neural networks,” *Neurocomputing*, vol. 24, no. 1, pp. 191–206, Feb. 1999, doi: 10.1016/S0925-2312(98)00102-7.
- [29] L. Rieger, P. Chormai, G. Montavon, L. K. Hansen, and K.-R. Müller, “Structuring Neural Networks for More Explainable Predictions,” in *Explainable and Interpretable Models in Computer Vision and Machine Learning*, H. J. Escalante, S. Escalera, I. Guyon, X. Baró, Y. Güçlütürk, U. Güçlü, and M. van Gerven, Eds., in The Springer Series on Challenges in Machine Learning. , Cham: Springer International Publishing, 2018, pp. 115–131. doi: 10.1007/978-3-319-98131-4_5.
- [30] M. Sari *et al.*, “Machine learning-based energy use prediction for the smart building energy management system,” *J. Inf. Technol. Constr. ITcon*, vol. 28, no. 33, pp. 622–645, Sep. 2023, doi: 10.36680/j.itcon.2023.033.
- [31] C. Li, Z. Ding, D. Zhao, J. Yi, and G. Zhang, “Building Energy Consumption Prediction: An Extreme Deep Learning Approach,” *Energies*, vol. 10, no. 10, Art. no. 10, Oct. 2017, doi: 10.3390/en10101525.
- [32] K. Li, R. J. Yang, D. Robinson, J. Ma, and Z. Ma, “An agglomerative hierarchical clustering-based strategy using Shared Nearest Neighbours and multiple dissimilarity measures to identify typical daily electricity usage profiles of university library buildings,” *Energy*, vol. 174, pp. 735–748, May 2019, doi: 10.1016/j.energy.2019.03.003.
- [33] K. Li, Z. Ma, D. Robinson, and J. Ma, “Identification of typical building daily electricity usage profiles using Gaussian mixture model-based clustering and hierarchical clustering,” *Appl. Energy*, vol. 231, pp. 331–342, Dec. 2018, doi: 10.1016/j.apenergy.2018.09.050.

- [34] W.-T. Wang, Y.-L. Wu, C.-Y. Tang, and M.-K. Hor, "Adaptive density-based spatial clustering of applications with noise (DBSCAN) according to data," in *2015 International Conference on Machine Learning and Cybernetics (ICMLC)*, Jul. 2015, pp. 445–451. doi: 10.1109/ICMLC.2015.7340962.
- [35] T. Räsänen, J. Ruuskanen, and M. Kolehmainen, "Reducing energy consumption by using self-organizing maps to create more personalized electricity use information," *Appl. Energy*, vol. 85, no. 9, pp. 830–840, Sep. 2008, doi: 10.1016/j.apenergy.2007.10.012.
- [36] A. E. Ezugwu *et al.*, "A comprehensive survey of clustering algorithms: State-of-the-art machine learning applications, taxonomy, challenges, and future research prospects," *Eng. Appl. Artif. Intell.*, vol. 110, p. 104743, Apr. 2022, doi: 10.1016/j.engappai.2022.104743.
- [37] A. K. Jain, "Data clustering: 50 years beyond K-means," *Pattern Recognit. Lett.*, vol. 31, no. 8, pp. 651–666, Jun. 2010, doi: 10.1016/j.patrec.2009.09.011.
- [38] B. Nepal, M. Yamaha, H. Sahashi, and A. Yokoe, "Analysis of Building Electricity Use Pattern Using K-Means Clustering Algorithm by Determination of Better Initial Centroids and Number of Clusters," *Energies*, vol. 12, no. 12, Art. no. 12, Jan. 2019, doi: 10.3390/en12122451.
- [39] A. Tureczek, P. S. Nielsen, and H. Madsen, "Electricity Consumption Clustering Using Smart Meter Data," *Energies*, vol. 11, no. 4, Art. no. 4, Apr. 2018, doi: 10.3390/en11040859.
- [40] M. Afaifia, K. A. Djiair, N. Bich-Ngoc, and J. Teller, "An energy consumption model for the Algerian residential building's stock, based on a triangular approach: Geographic Information System (GIS), regression analysis and hierarchical cluster analysis," *Sustain. Cities Soc.*, vol. 74, p. 103191, Nov. 2021, doi: 10.1016/j.scs.2021.103191.
- [41] P. Westermann, C. Deb, A. Schlueter, and R. Evins, "Unsupervised learning of energy signatures to identify the heating system and building type using smart meter data," *Appl. Energy*, vol. 264, p. 114715, Apr. 2020, doi: 10.1016/j.apenergy.2020.114715.
- [42] J. Clarke, *Energy Simulation in Building Design*, 2nd ed. London: Routledge, 2001. doi: 10.4324/9780080505640.
- [43] V. S. K. V. Harish and A. Kumar, "A review on modeling and simulation of building energy systems," *Renew. Sustain. Energy Rev.*, vol. 56, pp. 1272–1292, Apr. 2016, doi: 10.1016/j.rser.2015.12.040.
- [44] H. Zhao and F. Magoulès, "A review on the prediction of building energy consumption," *Renew. Sustain. Energy Rev.*, vol. 16, no. 6, pp. 3586–3592, Aug. 2012, doi: 10.1016/j.rser.2012.02.049.
- [45] F. C. Winkelmann *et al.*, "DOE-2 supplement: Version 2.1E," Lawrence Berkeley National Lab. (LBNL), Berkeley, CA (United States); Hirsch (James J.) and Associates, Camarillo, CA (United States), LBL-34947, Nov. 1993. doi: 10.2172/10147851.
- [46] R. Chargui and H. Sammouda, "Modeling of a residential house coupled with a dual source heat pump using TRNSYS software," *Energy Convers. Manag.*, vol. 81, pp. 384–399, May 2014, doi: 10.1016/j.enconman.2014.02.040.
- [47] I. Beausoleil-Morrison, M. Kummert, F. Macdonald, R. Jost, T. McDowell, and A. Ferguson, "Demonstration of the new ESP-r and TRNSYS co-simulator for modelling solar buildings," *Energy Procedia*, vol. 30, pp. 505–514, Jan. 2012, doi: 10.1016/j.egypro.2012.11.060.
- [48] D. B. Crawley, J. W. Hand, M. Kummert, and B. T. Griffith, "Contrasting the capabilities of building energy performance simulation programs," *Build. Environ.*, vol. 43, no. 4, pp. 661–673, Apr. 2008, doi: 10.1016/j.buildenv.2006.10.027.
- [49] M. Loyola, "Big data in building design: A review," *J. Inf. Technol. Constr.*, vol. 23, pp. 259–284, 2018.
- [50] A. Forrester, A. Sobester, and A. Keane, *Engineering Design via Surrogate Modelling: A Practical Guide*. John Wiley & Sons, 2008.
- [51] T. Simpson, V. Toropov, V. Balabanov, and F. Viana, "Design and Analysis of Computer Experiments in Multidisciplinary Design Optimization: A Review of How Far We Have Come - Or Not," in *12th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference*, American Institute of Aeronautics and Astronautics, 2012. doi: 10.2514/6.2008-5802.

- [52] L. Van Gelder, P. Das, H. Janssen, and S. Roels, “Comparative study of metamodeling techniques in building energy simulation: Guidelines for practitioners,” *Simul. Model. Pract. Theory*, vol. 49, pp. 245–257, Dec. 2014, doi: 10.1016/j.simpat.2014.10.004.
- [53] K. Bamdad, M. E. Cholette, and J. Bell, “Building energy optimization using surrogate model and active sampling,” *J. Build. Perform. Simul.*, vol. 13, no. 6, pp. 760–776, Nov. 2020, doi: 10.1080/19401493.2020.1821094.
- [54] B. Chegari, M. Tabaa, E. Simeu, F. Moutaouakkil, and H. Medromi, “An optimal surrogate-model-based approach to support comfortable and nearly zero energy buildings design,” *Energy*, vol. 248, p. 123584, Jun. 2022, doi: 10.1016/j.energy.2022.123584.
- [55] C. R. Lawrence, R. Richman, M. Kordjamshidi, and C. Skarupa, “Application of surrogate modelling to improve the thermal performance of single-family homes through archetype development,” *Energy Build.*, vol. 237, p. 110812, Apr. 2021, doi: 10.1016/j.enbuild.2021.110812.
- [56] P. Westermann and R. Evins, “Surrogate modelling for sustainable building design – A review,” *Energy Build.*, vol. 198, pp. 170–186, Sep. 2019, doi: 10.1016/j.enbuild.2019.05.057.
- [57] A. Bhosekar and M. Ierapetritou, “Advances in surrogate based modeling, feasibility analysis, and optimization: A review,” *Comput. Chem. Eng.*, vol. 108, pp. 250–267, Jan. 2018, doi: 10.1016/j.compchemeng.2017.09.017.
- [58] S. S. Garud, I. A. Karimi, and M. Kraft, “Design of computer experiments: A review,” *Comput. Chem. Eng.*, vol. 106, pp. 71–95, Nov. 2017, doi: 10.1016/j.compchemeng.2017.05.010.
- [59] R. Yondo, E. Andrés, and E. Valero, “A review on design of experiments and surrogate models in aircraft real-time and many-query aerodynamic analyses,” *Prog. Aerosp. Sci.*, vol. 96, pp. 23–61, Jan. 2018, doi: 10.1016/j.paerosci.2017.11.003.
- [60] H. Vallianos and C. Kapsis, “Surrogate models for evaluating HVAC retrofit options in multi-unit residential buildings,” *Energy Build.*, vol. 355, p. 117060, Mar. 2026, doi: 10.1016/j.enbuild.2026.117060.
- [61] B. Birdsell and R. Evins, “Surrogate Modeling performance for building design problems,” presented at the 2023 Building Simulation Conference, Sep. 2023. doi: 10.26868/25222708.2023.1376.
- [62] M. Beccali, G. Ciulla, V. Lo Brano, A. Galatioto, and M. Bonomolo, “Artificial neural network decision support tool for assessment of the energy performance and the refurbishment actions for the non-residential building stock in Southern Italy,” *Energy*, vol. 137, pp. 1201–1218, Oct. 2017, doi: 10.1016/j.energy.2017.05.200.
- [63] 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*, vol. 10, no. 12, p. 4718, Dec. 2018, doi: 10.3390/su10124718.
- [64] E. Thrampoulidis, G. Hug, and K. Orehounig, “Approximating optimal building retrofit solutions for large-scale retrofit analysis,” *Appl. Energy*, vol. 333, p. 120566, Mar. 2023, doi: 10.1016/j.apenergy.2022.120566.
- [65] H. Zhang, H. Feng, K. Hewage, and M. Arashpour, “Artificial Neural Network for Predicting Building Energy Performance: A Surrogate Energy Retrofits Decision Support Framework,” *Buildings*, vol. 12, no. 6, p. 829, Jun. 2022, doi: 10.3390/buildings12060829.
- [66] S. Ferreira, B. Gunay, A. Ashouri, and S. Shillinglaw, “Unsupervised learning of load signatures to estimate energy-related building features using surrogate modelling techniques,” *Build. Simul.*, vol. 16, no. 7, pp. 1273–1286, Jul. 2023, doi: 10.1007/s12273-023-1005-5.
- [67] A. Ali, R. Jayaraman, A. Mayyas, B. Alaifan, and E. Azar, “Machine learning as a surrogate to building performance simulation: Predicting energy consumption under different operational settings,” *Energy Build.*, vol. 286, p. 112940, May 2023, doi: 10.1016/j.enbuild.2023.112940.
- [68] E. Markarian, S. Qiblawi, S. Krishnan, A. Divakaran, A. Thomas, and E. Azar, “Evaluating the Capabilities of Machine Learning Surrogate Modeling to Support High-Resolution Building Performance Simulation,” in *Advances in Information Technology in Civil and Building*

- Engineering*, A. Francis, E. Miresco, and S. Melhado, Eds., Cham: Springer Nature Switzerland, 2025, pp. 169–180. doi: 10.1007/978-3-031-87364-5_14.
- [69] P. Govindarajan, F. P. Ortner, and J. M. Han, “Surrogate modeling: hourly energy performance prediction for residential precincts in tropical cities using synthetic data,” *Energy Build.*, vol. 347, p. 116366, Nov. 2025, doi: 10.1016/j.enbuild.2025.116366.
 - [70] L. Van Gelder, P. Das, H. Janssen, and S. Roels, “Comparative study of metamodeling techniques in building energy simulation: Guidelines for practitioners,” *Simul. Model. Pract. Theory*, vol. 49, pp. 245–257, Dec. 2014, doi: 10.1016/j.simpat.2014.10.004.
 - [71] J. S. Hygh, J. F. DeCarolís, D. B. Hill, and S. Ranji Ranjithan, “Multivariate regression as an energy assessment tool in early building design,” *Build. Environ.*, vol. 57, pp. 165–175, Nov. 2012, doi: 10.1016/j.buildenv.2012.04.021.
 - [72] Y. Liang, Y. Pan, X. Yuan, W. Jia, and Z. Huang, “Surrogate modeling for long-term and high-resolution prediction of building thermal load with a metric-optimized KNN algorithm,” *Energy Built Environ.*, vol. 4, no. 6, pp. 709–724, Dec. 2023, doi: 10.1016/j.enbenv.2022.06.008.
 - [73] X. Liu, Y. Ding, H. Tang, and F. Xiao, “A data mining-based framework for the identification of daily electricity usage patterns and anomaly detection in building electricity consumption data,” *Energy Build.*, vol. 231, p. 110601, Jan. 2021, doi: 10.1016/j.enbuild.2020.110601.
 - [74] B. Eisenhower, Z. O’Neill, V. A. Fonoberov, and I. Mezić, “Uncertainty and sensitivity decomposition of building energy models,” *J. Build. Perform. Simul.*, vol. 5, no. 3, pp. 171–184, May 2012, doi: 10.1080/19401493.2010.549964.
 - [75] I. Macdonald and P. Strachan, “Practical application of uncertainty analysis,” *Energy Build.*, vol. 33, no. 3, pp. 219–227, Feb. 2001, doi: 10.1016/S0378-7788(00)00085-2.
 - [76] L. Rivalin, P. Stabat, D. Marchio, M. Caciolo, and F. Hopquin, “A comparison of methods for uncertainty and sensitivity analysis applied to the energy performance of new commercial buildings,” *Energy Build.*, vol. 166, pp. 489–504, May 2018, doi: 10.1016/j.enbuild.2018.02.021.
 - [77] J. Carreras *et al.*, “Systematic approach for the life cycle multi-objective optimization of buildings combining objective reduction and surrogate modeling,” *Energy Build.*, vol. 130, pp. 506–518, Oct. 2016, doi: 10.1016/j.enbuild.2016.07.062.
 - [78] E. Gengembre, B. Ladevie, O. Fudym, and A. Thuillier, “A Kriging constrained efficient global optimization approach applied to low-energy building design problems,” *Inverse Probl. Sci. Eng.*, vol. 20, no. 7, pp. 1101–1114, Oct. 2012, doi: 10.1080/17415977.2012.727084.
 - [79] C.-H. Park and C. S. Park, “Limitations and issues of conventional artificial neural network-based surrogate models for building energy retrofit,” *J. Build. Perform. Simul.*, vol. 17, no. 3, pp. 361–370, May 2024, doi: 10.1080/19401493.2023.2282078.
 - [80] B. Gunay, B. W. Hobson, D. Darwazeh, and J. Bursill, “Estimating energy savings from HVAC controls fault correction through inverse greybox model-based virtual metering,” *Energy Build.*, vol. 282, p. 112806, Mar. 2023, doi: 10.1016/j.enbuild.2023.112806.
 - [81] G. James, D. Witten, T. Hastie, R. Tibshirani, and J. Taylor, *An Introduction to Statistical Learning: with Applications in Python*. in Springer Texts in Statistics. Cham: Springer International Publishing, 2023. doi: 10.1007/978-3-031-38747-0.
 - [82] P. Semler and M. Weiser, “Adaptive Gaussian process regression for efficient building of surrogate models in inverse problems,” *Inverse Probl.*, vol. 39, no. 12, p. 125003, Oct. 2023, doi: 10.1088/1361-6420/ad0028.
 - [83] P. Westermann and R. Evins, “Using Bayesian deep learning approaches for uncertainty-aware building energy surrogate models,” *Energy AI*, vol. 3, p. 100039, Mar. 2021, doi: 10.1016/j.egyai.2020.100039.
 - [84] D. Hou and R. Evins, “A protocol for developing and evaluating neural network-based surrogate models and its application to building energy prediction,” *Renew. Sustain. Energy Rev.*, vol. 193, p. 114283, Apr. 2024, doi: 10.1016/j.rser.2024.114283.

- [85] 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.
- [86] M. Shi, L. Lv, W. Sun, and X. Song, "A multi-fidelity surrogate model based on support vector regression," *Struct. Multidiscip. Optim.*, vol. 61, no. 6, pp. 2363–2375, Jun. 2020, doi: 10.1007/s00158-020-02522-6.
- [87] A. Forrester, A. Sóbester, and A. Keane, *Engineering Design via Surrogate Modelling: A Practical Guide*. John Wiley & Sons, 2008.
- [88] 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.
- [89] S. Ferreira, B. Gunay, and A. Wills, "Artificial Neural Network-Based Surrogate Models to Predict Building Features from Load Signatures," *ASHRAE Trans.*, vol. 129, pp. 831–839, 2023.
- [90] T.-Y. Kim and S.-B. Cho, "Predicting residential energy consumption using CNN-LSTM neural networks," *Energy*, vol. 182, pp. 72–81, Sep. 2019, doi: 10.1016/j.energy.2019.05.230.
- [91] Z. Wang, T. Hong, and M. A. Piette, "Building thermal load prediction through shallow machine learning and deep learning," *Appl. Energy*, vol. 263, p. 114683, Apr. 2020, doi: 10.1016/j.apenergy.2020.114683.
- [92] M. Li *et al.*, "A deep learning convolutional neural network and multi-layer perceptron hybrid fusion model for predicting the mechanical properties of carbon fiber," *Mater. Des.*, vol. 227, p. 111760, Mar. 2023, doi: 10.1016/j.matdes.2023.111760.
- [93] J. Fañanás-Anaya, G. López-Nicolás, and C. Sagüés, "Dynamical system simulation with attention and recurrent neural networks," *Neural Comput. Appl.*, vol. 37, no. 4, pp. 2711–2731, Feb. 2025, doi: 10.1007/s00521-024-10732-y.
- [94] T. (中村太一) Nakamura, K. (深見開) Fukami, K. (長谷川一登) Hasegawa, Y. (難波江佑介) Nabae, and K. (深淵康二) Fukagata, "Convolutional neural network and long short-term memory based reduced order surrogate for minimal turbulent channel flow," *Phys. Fluids*, vol. 33, no. 2, p. 025116, Feb. 2021, doi: 10.1063/5.0039845.
- [95] C. Fan, J. Wang, W. Gang, and S. Li, "Assessment of deep recurrent neural network-based strategies for short-term building energy predictions," *Appl. Energy*, vol. 236, pp. 700–710, Feb. 2019, doi: 10.1016/j.apenergy.2018.12.004.
- [96] F. Madaeni, K. Chokmani, R. Lhissou, S. Homayouni, Y. Gauthier, and S. Tolszczuk-Leclerc, "Convolutional neural network and long short-term memory models for ice-jam predictions," *The Cryosphere*, vol. 16, no. 4, pp. 1447–1468, Apr. 2022, doi: 10.5194/tc-16-1447-2022.
- [97] Z. Shi, Y. Hu, G. Mo, and J. Wu, "Attention-based CNN-LSTM and XGBoost hybrid model for stock prediction," arXiv.org. Accessed: Feb. 07, 2026. [Online]. Available: <https://arxiv.org/abs/2204.02623v2>
- [98] G. Li, W. Tian, H. Zhang, and B. Chen, "Building Energy Models at Different Time Scales Based on Multi-Output Machine Learning," *Buildings*, vol. 12, no. 12, p. 2109, Dec. 2022, doi: 10.3390/buildings12122109.
- [99] Z. Feng *et al.*, "A hybrid CNN-transformer surrogate model for the multi-objective robust optimization of geological carbon sequestration," *Adv. Water Resour.*, vol. 196, p. 104897, Feb. 2025, doi: 10.1016/j.advwatres.2025.104897.
- [100] J. H. Park, J.-B. Kim, S. Lee, J. Kang, and D. Mun, "Hybrid MLP-CNN-based ground sink susceptibility prediction in urban area using underground pipe map," *Reliab. Eng. Syst. Saf.*, vol. 245, p. 110031, May 2024, doi: 10.1016/j.res.2024.110031.
- [101] A. Zhao, L. Mi, X. Xue, J. Xi, and Y. Jiao, "Heating load prediction of residential district using hybrid model based on CNN," *Energy Build.*, vol. 266, p. 112122, Jul. 2022, doi: 10.1016/j.enbuild.2022.112122.

- [102] S. Du, T. Li, Y. Yang, and S.-J. Horng, "Multivariate time series forecasting via attention-based encoder–decoder framework," *Neurocomputing*, vol. 388, pp. 269–279, May 2020, doi: 10.1016/j.neucom.2019.12.118.
- [103] R. Chandra, S. Goyal, and R. Gupta, "Evaluation of Deep Learning Models for Multi-Step Ahead Time Series Prediction," *IEEE Access*, vol. 9, pp. 83105–83123, 2021, doi: 10.1109/ACCESS.2021.3085085.
- [104] Y. Yang, Q. Fu, J. Chen, Y. Lu, Y. Wang, and H. Wu, "Temporal feature decomposition fusion network for building energy multi-step prediction," *J. Build. Eng.*, vol. 94, p. 110034, Oct. 2024, doi: 10.1016/j.job.2024.110034.
- [105] S. S. Ali, V. S. Yadav, B. Nouri, and A. Ghani, "Hybrid Surrogate Modeling Approach for Data Reduction and Design Space Exploration of Turbine Blades," *J. Eng. Gas Turbines Power*, vol. 147, no. 071004, Dec. 2024, doi: 10.1115/1.4066998.
- [106] Q. Liu, F. Lanfermann, T. Rodemann, M. Olhofer, and Y. Jin, "Surrogate-Assisted Many-Objective Optimization of Building Energy Management," *IEEE Comput. Intell. Mag.*, vol. 18, no. 4, pp. 14–28, Nov. 2023, doi: 10.1109/MCI.2023.3304073.
- [107] S. Ferreira, B. Gunay, A. Wills, and F. Rizvi, "A neural network-based surrogate model to predict building features from heating and cooling load signatures," *J. Build. Perform. Simul.*, vol. 17, no. 5, pp. 631–654, Sep. 2024, doi: 10.1080/19401493.2024.2375304.
- [108] C. J. Hopfe and J. L. M. Hensen, "Uncertainty analysis in building performance simulation for design support," *Energy Build.*, vol. 43, no. 10, pp. 2798–2805, Oct. 2011, doi: 10.1016/j.enbuild.2011.06.034.
- [109] T. Hong, J. Langevin, and K. Sun, "Building simulation: Ten challenges," *Build. Simul.*, vol. 11, no. 5, pp. 871–898, Oct. 2018, doi: 10.1007/s12273-018-0444-x.
- [110] S. K. Kwak and J. H. Kim, "Statistical data preparation: management of missing values and outliers," *Korean J. Anesthesiol.*, vol. 70, no. 4, pp. 407–411, Jul. 2017, doi: 10.4097/kjae.2017.70.4.407.
- [111] S. G. K. Patro and K. K. Sahu, "Normalization: A Preprocessing Stage," Mar. 19, 2015, *arXiv*: arXiv:1503.06462. doi: 10.48550/arXiv.1503.06462.
- [112] V. N. G. Raju, K. P. Lakshmi, V. M. Jain, A. Kalidindi, and V. Padma, "Study the Influence of Normalization/Transformation process on the Accuracy of Supervised Classification," in *2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT)*, Aug. 2020, pp. 729–735. doi: 10.1109/ICSSIT48917.2020.9214160.
- [113] H. P. Vinutha, B. Poornima, and B. M. Sagar, "Detection of Outliers Using Interquartile Range Technique from Intrusion Dataset," in *Information and Decision Sciences*, S. C. Satapathy, J. M. R. S. Tavares, V. Bhateja, and J. R. Mohanty, Eds., Singapore: Springer, 2018, pp. 511–518. doi: 10.1007/978-981-10-7563-6_53.
- [114] B. Rosner, "Percentage Points for a Generalized ESD Many-Outlier Procedure," *Technometrics*, vol. 25, no. 2, pp. 165–172, May 1983, doi: 10.1080/00401706.1983.10487848.
- [115] L. Czétány *et al.*, "Development of electricity consumption profiles of residential buildings based on smart meter data clustering," *Energy Build.*, vol. 252, p. 111376, Dec. 2021, doi: 10.1016/j.enbuild.2021.111376.
- [116] M. Cui, "Introduction to the K-Means Clustering Algorithm Based on the Elbow Method," 2020, doi: 10.23977/accaf.2020.010102.
- [117] P. J. Rousseeuw, "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis," *J. Comput. Appl. Math.*, vol. 20, pp. 53–65, Nov. 1987, doi: 10.1016/0377-0427(87)90125-7.
- [118] Z. Ding, Z. Wang, T. Hu, and H. Wang, "A Comprehensive Study on Integrating Clustering with Regression for Short-Term Forecasting of Building Energy Consumption: Case Study of a Green Building," *Buildings*, vol. 12, no. 10, p. 1701, Oct. 2022, doi: 10.3390/buildings12101701.
- [119] A. El Saied, C. Maalouf, T. Bejat, and E. Wurtz, "Thermal bridges numerical study in building retrofitting and ground model coupling with Whole building energy software 'EnergyPlus,'" *Energy Build.*, vol. 252, p. 111376, Dec. 2021, doi: 10.1016/j.enbuild.2021.111376.

- presented at the Building Simulation 2021, in Building Simulation, vol. 17. IBPSA, 2021, pp. 2211–2218. doi: 10.26868/25222708.2021.30967.
- [120] “EnergyPlus Testing with ANSI/ASHRAE Standard,” 2004.
 - [121] E. Erdogan, “Comparative studies of building performance simulation tools (EnergyPlus, IES-VE, IDA-ICE),” Oct. 2020, Accessed: Jun. 30, 2024. [Online]. Available: <https://www.politesi.polimi.it/handle/10589/166623>
 - [122] A. Roth *et al.*, “Scripting Frameworks for Enhancing EnergyPlus Modeling Productivity,” presented at the 2018 Building Performance Analysis Conference and SimBuild, in SimBuild Conference, vol. 8. ASHRAE/IBPSA-USA, 2018, pp. 312–319. Accessed: Jun. 30, 2024. [Online]. Available: https://publications.ibpsa.org/conference/paper/?id=simbuild2018_C043
 - [123] P. C. Tabares-Velasco, C. Christensen, and M. Bianchi, “Verification and validation of EnergyPlus phase change material model for opaque wall assemblies,” *Build. Environ.*, vol. 54, pp. 186–196, Aug. 2012, doi: 10.1016/j.buildenv.2012.02.019.
 - [124] A. Saltelli *et al.*, *Global Sensitivity Analysis: The Primer*. John Wiley & Sons, 2008.
 - [125] W. Tian, “A review of sensitivity analysis methods in building energy analysis,” *Renew. Sustain. Energy Rev.*, vol. 20, pp. 411–419, Apr. 2013, doi: 10.1016/j.rser.2012.12.014.
 - [126] “One-Factor-at-a-Time versus Designed Experiments: The American Statistician: Vol 53, No 2.” Accessed: Feb. 02, 2026. [Online]. Available: <https://www.tandfonline.com/doi/abs/10.1080/00031305.1999.10474445>
 - [127] P. Heiselberg, H. Brohus, A. Hesselholt, H. Rasmussen, E. Seinre, and S. Thomas, “Application of sensitivity analysis in design of sustainable buildings,” *Renew. Energy*, vol. 34, no. 9, pp. 2030–2036, Sep. 2009, doi: 10.1016/j.renene.2009.02.016.
 - [128] K. Menberg, Y. Heo, and R. Choudhary, “Sensitivity analysis methods for building energy models: Comparing computational costs and extractable information,” *Energy Build.*, vol. 133, pp. 433–445, Dec. 2016, doi: 10.1016/j.enbuild.2016.10.005.
 - [129] A. Altmann, L. Toloşi, O. Sander, and T. Lengauer, “Permutation importance: a corrected feature importance measure,” Accessed: Feb. 02, 2026. [Online]. Available: <https://dx.doi.org/10.1093/bioinformatics/btq134>
 - [130] M. Mikhail, D. Mather, P. Parker, and K. Kapsis, “Net-positive office commissioning and performance gap assessment: Empirical insights,” *Energy Build.*, vol. 279, p. 112717, Jan. 2023, doi: 10.1016/j.enbuild.2022.112717.
 - [131] VCT Group, “evolv1,” VCTGroup.com. Accessed: Aug. 17, 2024. [Online]. Available: <https://vctgroup.com/evolv1/>
 - [132] Y. Wei *et al.*, “A review of data-driven approaches for prediction and classification of building energy consumption,” *Renew. Sustain. Energy Rev.*, vol. 82, pp. 1027–1047, Feb. 2018, doi: 10.1016/j.rser.2017.09.108.
 - [133] H. Fu, J.-C. Baltazar, and D. E. Claridge, “Review of developments in whole-building statistical energy consumption models for commercial buildings,” *Renew. Sustain. Energy Rev.*, vol. 147, p. 111248, Sep. 2021, doi: 10.1016/j.rser.2021.111248.
 - [134] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, “Optuna: A Next-generation Hyperparameter Optimization Framework,” in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, in KDD ’19. New York, NY, USA: Association for Computing Machinery, Jul. 2019, pp. 2623–2631. doi: 10.1145/3292500.3330701.
 - [135] “ASHRAE Guideline 14-2023 -- Measurement of Energy, Demand and Water Savings | ASHRAE Store.” Accessed: Apr. 29, 2026. [Online]. Available: https://store.accuristech.com/standards/ashrae-guideline-14-2023-measurement-of-energy-demand-and-water-savings?product_id=2569793
 - [136] “ANSI/ASHRAE Addendum a to ANSI/ASHRAE Standard 140-2017”.

- [137] P. Im, J. Joe, Y. Bae, and J. R. New, “Empirical validation of building energy modeling for multi-zones commercial buildings in cooling season,” *Appl. Energy*, vol. 261, p. 114374, Mar. 2020, doi: 10.1016/j.apenergy.2019.114374.
- [138] Z. Xin *et al.*, “Update of ASHRAE Standard 140/BESTEST thermal fabric test cases,” presented at the 2021 Building Simulation Conference, Sep. 2021. doi: 10.26868/25222708.2021.30365.
- [139] T. N. Aynur, “Variable refrigerant flow systems: A review,” *Energy Build.*, vol. 42, no. 7, pp. 1106–1112, Jul. 2010, doi: 10.1016/j.enbuild.2010.01.024.
- [140] “High Performance Wall Assemblies.” Ontario Association of Architects. [Online]. Available: https://oaa.on.ca/Assets/Common/Shared_Documents/Tools-Resources/05.Documents/High-Performance-Wall-Assemblies.pdf
- [141] K. Voss and E. Musall, *Net zero energy buildings: International projects of carbon neutrality in buildings*. Walter de Gruyter, 2012.
- [142] F. Goia, “Search for the optimal window-to-wall ratio in office buildings in different European climates and the implications on total energy saving potential,” *Sol. Energy*, vol. 132, pp. 467–492, Jul. 2016, doi: 10.1016/j.solener.2016.03.031.
- [143] R. Azari and N. Badri, “Life Cycle Assessment as a Research Methodology for Estimating the Environmental Impacts of Buildings,” in *Research Methods in Building Science and Technology*, R. Azari and H. Rashed-Ali, Eds., Cham: Springer International Publishing, 2021, pp. 151–173. doi: 10.1007/978-3-030-73692-7_8.
- [144] V. Mykytyak, “The impact of window-to-wall ratio on energy intensity of existing office buildings in Ontario and Quebec,” thesis, Toronto Metropolitan University, 2021. doi: 10.32920/ryerson.14665374.v1.
- [145] L. C. Ng, A. K. Persily, and S. J. Emmerich, “Improving infiltration modeling in commercial building energy models,” *Energy Build.*, vol. 88, pp. 316–323, Feb. 2015, doi: 10.1016/j.enbuild.2014.11.078.
- [146] K. Gowri, D. W. Winiarski, and R. E. Jarnagin, “Infiltration modeling guidelines for commercial building energy analysis,” Pacific Northwest National Lab. (PNNL), Richland, WA (United States), PNNL-18898, Sep. 2009. doi: 10.2172/968203.

Appendices

Appendix Table List

Appendix Table 1- Occupancy, Lighting and Equipment schedule for office buildings based on NECB (adopted from [10]).....	144
Appendix Table 2- HVC setpoint schedule for parametrical numerical model based on high-performance case study building.....	144
Appendix Table 3- Versions of Libraries needed for Python codes	146

Appendix-A: Schedule Input Values

Appendix Table 1- Occupancy, Lighting and Equipment schedule for office buildings based on NECB (adopted from [10])

Day	Time of the day																							
	1 AM	2 AM	3 AM	4 AM	5 AM	6 AM	7 AM	8 AM	9 AM	10 AM	11 AM	12 PM	1 PM	2 PM	3 PM	4 PM	5 PM	6 PM	7 PM	8 PM	9 PM	10 PM	11 PM	12 AM
Occupant, Fraction occupied																								
Mon-Fri	0	0	0	0	0	0	0.1	0.7	0.9	0.9	0.9	0.5	0.5	0.9	0.9	0.9	0.7	0.3	0.1	0.1	0.1	0.1	0	0
Sat, Sun	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Lighting, Fraction "ON"																								
Mon-Thu	0.05	0.05	0.05	0.05	0.05	0.05	0.3	0.8	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.8	0.5	0.3	0.3	0.1	0.1	0.05	0.05
Sat, Sun	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05
Receptacle Equipment, Fraction of load																								
Mon-Fri	0.2	0.2	0.2	0.2	0.2	0.2	0.3	0.8	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.5	0.3	0.3	0.2	0.2	0.2	0.2
Sat, Sun	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2

Appendix Table 2- HVC setpoint schedule for parametrical numerical model based on high-performance case study building

Day	Time of the day																							
	1 AM	2 AM	3 AM	4 AM	5 AM	6 AM	7 AM	8 AM	9 AM	10 AM	11 AM	12 PM	1 PM	2 PM	3 PM	4 PM	5 PM	6 PM	7 PM	8 PM	9 PM	10 PM	11 PM	12
Winter Heating Hourly Setpoint Schedule (°C)																								
Mon-Fri	21	21	21	21	22.5	22.5	22.5	22.5	22.5	22.5	22.5	22.5	22.5	22.5	22.5	22.5	22.5	22.5	21	21	21	21	21	21
Sat, Sun	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21
Winter Cooling Hourly Setpoint Schedule(°C)																								
Mon-Thu	26	26	26	26	24.5	24.5	24.5	24.5	24.5	24.5	24.5	24.5	24.5	24.5	24.5	24.5	24.5	24.5	26	26	26	26	26	26
Fri	26	26	26	26	24.5	24.5	24.5	24.5	24.5	24.5	24.5	24.5	24.5	24.5	24.5	24.5	24.5	24.5	27	27	27	27	27	27
Sat, Sun	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27
Summer Heating Hourly Setpoint Schedule(°C)																								
Mon-Fri	18	18	18	18	21.5	21.5	21.5	21.5	21.5	21.5	21.5	21.5	21.5	21.5	21.5	21.5	21.5	21.5	18	18	18	18	18	18
Sat, Sun	18	18	18	18	18	18	18	18	18	18	18	18	18	18	18	18	18	18	18	18	18	18	18	18
Summer Cooling Hourly Setpoint Schedule(°C)																								
Mon-Thu	26	26	26	26	23.5	23.5	23.5	23.5	23.5	23.5	23.5	23.5	23.5	23.5	23.5	23.5	23.5	23.5	26	26	26	26	26	26
Fri	26	26	26	26	23.5	23.5	23.5	23.5	23.5	23.5	23.5	23.5	23.5	23.5	23.5	23.5	23.5	23.5	27	27	27	27	27	27
Sat, Sun	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27	27

Appendix-B: Python Code Requirements

Based on the code developed for case study investigation, parametric model development, training and validation of the ensemble surrogate model, sensitivity analysis, clustering, and visualization, the following Python libraries were used.

Core Scientific & Data Handling

- **NumPy** – numerical operations and array handling
- **Pandas** – data storage, CSV handling, and tabular operations
- **SciPy** – statistical analysis, including Pearson and Spearman correlations
- **Machine Learning & Deep Learning**
- **PyTorch** – development and training of ANN, CNN, and hybrid ANN–CNN models
- **scikit-learn** –
 - permutation importance
 - K-means clustering
 - evaluation metrics (RMSE, R^2)
 - preprocessing utilities

Optimization

- **Optuna** – hyperparameter optimization for machine learning models

Sensitivity Analysis

- **SALib (conceptually)** – Morris method logic for sensitivity analysis
(Note: the Morris method was implemented manually rather than using SALib's API.)

Visualization

- **Matplotlib** – plotting loss curves, scatter plots, and general figures
- **Seaborn** – heatmaps and statistical visualizations
- **Calmap** – calendar-based heatmap visualization used to illustrate temporal patterns in daily energy consumption

- **h5py** – HDF5 storage for large hourly datasets
- **json** – saving sensitivity analysis outputs and experiment metadata
- **os / pathlib** – file and directory handling

Recommended Versions

The following versions of the libraries presented in Appendix Table 3 are recommended for running the codes generated for this study.

Appendix Table 3- Versions of Libraries needed for Python codes

Library	Version
Python	≥ 3.10
NumPy	≥ 1.23
Pandas	≥ 1.5
SciPy	≥ 1.9
PyTorch	≥ 2.0
scikit-learn	≥ 1.2
Optuna	≥ 3.0
Matplotlib	≥ 3.7
Seaborn	≥ 0.12
calmap	$\geq 0.0.11$
h5py	≥ 3.8

All simulations, data processing, and machine learning workflows were implemented in **Python (version 3.10+)**. Core numerical and data handling operations were performed using **NumPy** and **Pandas**, while model development and training were carried out using **PyTorch**. Hyperparameter optimization was conducted using **Optuna**, and model evaluation, clustering, and permutation-based sensitivity analysis were implemented using **scikit-learn**. Statistical analyses relied on **SciPy**, and visualizations were generated using **Matplotlib**, **Seaborn**, and **Calmap** for calendar-based time-series visualization. Large hourly datasets were stored using the **HDF5 format via h5py**, while experiment outputs and metadata were managed using **JSON** files and standard Python file-handling utilities. All libraries were used in stable, widely adopted versions consistent with current best practices in building energy modelling and machine learning research.

Appendix-C: Python Notebooks and GitHub links

The following Python scripts and notebooks were developed and used throughout this research for data analysis, clustering, numerical model development and data synthetic generation, surrogate model development, and sensitivity analysis. The link for accessing each code through GitHub has been presented after explanations for each code:

1- **Evolv Hourly Energy Analysis.ipynb:**

This notebook performs the initial energy behavior analysis of the case study office building. The analysis focuses on the total building electricity consumption to identify daily and seasonal patterns in energy use.

[Link](#)

2- **Evolv_1 clustering divided loads and index calculations.ipynb:**

This notebook analyzes individual energy loads and their contribution to total building energy consumption. It also calculates key building energy performance indices used throughout the study.

[Link](#)

3- **FC_ ANN K-fold for Surrogate model_R01.ipynb:**

This script implements the training and validation workflow for the surrogate model using a fully connected artificial neural network (ANN). Model performance is evaluated using K-fold cross-validation.

[Link](#)

4- **CNN K-fold for Surrogate model.ipynb:**

his notebook implements the surrogate model using a convolutional neural network (CNN) architecture and evaluates its predictive performance through K-fold cross-validation.

[Link](#)

5- **Hybrid ANN-CNN K-fold for Surrogate model_R01.ipynb:**

This notebook presents the development of the hybrid ANN–CNN surrogate model. It includes model training, evaluation, and sensitivity analysis, and generates the final standalone trained model used in the study.

[Link](#)