Data-Driven Methodologies and Advanced Machine Learning Models for Predicting, Evaluating, and Recommending Energy Load Measures and Analysis in District Area: Application of Artificial Intelligence

by

Amir SHAHCHERAGHIAN

MANUSCRIPT BASED PRESENTED TO ÉCOLE DE TECHNOLOGIE SUPÉRIEURE IN PARTIAL FULFILLEMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTORATE IN ENGINEERING
PH. D.

MONTREAL, JUNE 16, 2025

ÉCOLE DE TECHNOLOGIE SUPÉRIEURE
UNIVERSITÉ DU QUÉBEC

BOARD OF EXAMINERS

THIS THESIS HAS BEEN EVALUATED

BY THE FOLLOWING BOARD OF EXAMINERS

M. Adrian Ilinca, Thesis Supervisor
Department of Mechanical Engineering, École de technologie supérieure

Mme Danielle Monfet, President of the Board of Examiners
Department of Construction Engineering, École de technologie supérieure

M. Stanislaw Kajl, Member of the Jury
Department of Mechanical Engineering, École de technologie supérieure

M. Yves Gagnon, External Examiner
Faculty of Engineering, Université de Moncton

THIS THESIS WAS PRESENTED AND DEFENDED

IN THE PRESENCE OF A BOARD OF EXAMINERS AND PUBLIC

ON MAY 7, 2025

AT ÉCOLE DE TECHNOLOGIE SUPÉRIEURE

MÉTHODOLOGIES BASÉES SUR LES DONNÉES ET MODÈLES AVANCÉS D'APPRENTISSAGE MACHINE POUR LA PRÉDICTION, L'ÉVALUATION ET LA RECOMMANDATION DE MESURES DE CHARGE ÉNERGÉTIQUE ET D'ANALYSE DANS LES ZONES DE DISTRICT.

Amir SHAHCHERAGHIAN

RESUME

Plusieurs méthodes de modélisation des systèmes énergétiques des districts ont été proposées comme solutions potentielles aux défis posés par la Modélisation des systèmes énergétiques des bâtiments et des districts. Une approche prometteuse consiste à utiliser des modèles à boîte blanche, qui reposent sur des équations physiques détaillées et des paramètres pour simuler avec précision les flux d'énergie et la dynamique des systèmes. Ces modèles offrent une grande précision et une compréhension approfondie des processus physiques au sein des bâtiments et des districts. Cependant, les modèles de type boîte blanche peuvent être intensifs en calcul et nécessitent souvent des données d'entrée très détaillées. D'autres approches, telles que les modèles de type boîte noire et les modèles hybrides, sont également explorées pour équilibrer la précision avec l'efficacité computationnelle, en particulier lorsque la disponibilité des données ou la complexité des systèmes limite la praticité des modèles à boîte blanche.

Un autre méthode consiste à utiliser des modèles à boîte noire, y compris des algorithmes d'apprentissage automatique et des réseaux neuronaux artificiels. Les modèles à boîte noire se concentrent sur la reconnaissance des motifs et les relations statistiques dans les données plutôt que sur les principes physiques. Ils sont particulièrement utiles lorsque des données historiques abondantes sont disponibles, permettant de modéliser efficacement des relations complexes et non linéaires sans nécessiter de connaissances détaillées du système. Cependant, ces modèles manquent souvent d'interprétabilité, ce qui rend difficile la compréhension des processus sous-jacents et des relations causales. Malgré cela, les modèles à boîte noire peuvent offrir une grande précision prédictive et sont moins gourmands en ressources computationnelles, ce qui les rend adaptés à la modélisation énergétique à grande échelle dans les districts, où des approximations rapides sont nécessaires.

Cette étude se concentre sur l'utilisation d'applications d'apprentissage machine et de réseaux neuronaux profonds pour prédire et évaluer les mesures de demande énergétique et effectuer des analyses détaillées. Plus précisément, elle vise à développer des modèles capables de prévoir la demande énergétique, d'évaluer la distribution des charges et d'identifier les schémas de consommation dans diverses zones. L'étude inclut l'application de méthodes de régression, de classification et de regroupement (clustering), ainsi que des réseaux neuronaux profonds, pour capturer efficacement les tendances de consommation d'énergie linéaires et non linéaires. En utilisant ces techniques, l'étude cherche à générer des prévisions précises des charges énergétiques et à découvrir des informations sur les facteurs influençant la demande énergétique dans des environnements urbains complexes.

La recherche s'applique à des études de cas concernant le campus de l'Université de la Colombie-Britannique (UBC) à Vancouver et la ville de Stockholm en Suède. Ces zones présentent des caractéristiques d'utilisation énergétique uniques en raison de leurs climats, infrastructures et densités de population différentes, ce qui en fait des cas idéales pour des analyses comparatives. Les objectifs de l'étude incluent l'amélioration de l'efficacité des systèmes de distribution d'énergie, l'identification des périodes de demande de pointe et le soutien à la planification énergétique durable pour de grandes zones de district. De plus, les informations recueillies peuvent éclairer les pratiques de gestion énergétique à l'échelle des districts et faciliter l'intégration des sources d'énergie renouvelable, contribuant ainsi au développement de systèmes énergétiques urbains plus intelligents et résilients.

Cette thèse développe et évalue des méthodologies basées sur les données ainsi que des modèles avancés d'apprentissage automatique (ML) pour la prédiction, l'analyse et la recommandation de mesures de charge énergétique dans les systèmes énergétiques urbains à l'échelle des quartiers. Elle compare des outils de modélisation de type « boîte blanche », « boîte noire » et hybrides, en mettant l'accent sur leur applicabilité dans les pratiques de conseil en énergie. La recherche se concentre sur deux études de cas principales : le campus de l'Université de la Colombie-Britannique (UBC) et la ville de Stockholm.

Pour UBC, divers modèles de ML et d'apprentissage profond notamment les arbres de décision, la régression par vecteurs de support et les réseaux de neurones artificiels ont été appliqués pour prévoir la consommation d'électricité, d'eau chaude et de gaz. Ces modèles ont atteint une précision prédictive élevée (par exemple, R^2 allant jusqu'à 0,94, faible MAE), et leurs performances ont été validées à l'aide d'une division structurée des données en ensembles d'entraînement, de test et de validation.

Pour Stockholm, des méthodes de regroupement spatial (K-moyennes, regroupement agglomératif) ont été utilisées pour cartographier la demande de chaleur et attribuer les sources de chaleur résiduelle. Le coût actualisé de la chaleur (LCOH) a été calculé par groupe afin de soutenir la planification des infrastructures. Les résultats démontrent l'efficacité de la modélisation axée sur les données tant pour la prévision temporelle que pour l'optimisation énergétique spatiale. Les méthodologies proposées sont généralisables et peuvent guider des stratégies énergétiques urbaines évolutives et à faible émission de carbone.

Motsclés : Modélisation énergétique des districts, Modèles à boîte blanche, Modèles à boîte noire, Apprentissage automatique, Prédiction des charges énergétiques

DATA-DRIVEN METHODOLOGIES AND ADVANCED MACHINE LEARNING MODELS FOR PREDICTING, EVALUATING, AND RECOMMENDING ENERGY LOAD MEASURES AND ANALYSIS IN DISTRICT AREAS. APPLICATION OF ARTIFICIAL INTELLIGENCE.

Amir SHAHCHERAGHIAN

ABSTRACT

Several methods for district energy modeling have been proposed as potential solutions to the challenges posed by building and district energy modeling. One promising approach involves using white box models, which rely on detailed physical equations and parameters to simulate energy flows and system dynamics accurately. These models provide high accuracy and insight into the physical processes within buildings and districts. However, white box models can be computationally intensive and often require extensive data inputs. Other approaches, such as black box, are also explored to balance accuracy with computational efficiency, especially when data availability or system complexity limits the practicality of white box models.

Another method involves using black box models, including machine learning algorithms and artificial neural networks. Black box models focus on pattern recognition and statistical relationships within data rather than physical principles. They are particularly useful when extensive historical data is available, enabling them to effectively model complex, nonlinear relationships without detailed system knowledge. However, these models often lack interpretability, challenging the understanding of underlying processes and causal relationships. Despite this, black box models can offer high predictive accuracy and are computationally less intensive, making them suitable for large-scale district energy modeling where quick approximations are necessary.

This study uses machine learning and deep neural network applications to predict and evaluate energy load metrics and conduct detailed district-level analyses. Specifically, it aims to develop models to forecast energy demand, assess load distribution, and identify consumption patterns across various zones. The study includes the application of regression, classification, and clustering methods alongside deep neural networks to effectively capture both linear and nonlinear energy consumption trends. By employing these techniques, the study seeks to generate accurate energy load predictions and uncover insights into the factors driving energy demand in complex urban environments.

The research is applied to case studies involving the University of British Columbia (UBC) campus in Vancouver and the city of Stockholm in Sweden. These areas present unique energy use characteristics due to differing climates, infrastructure, and population densities, making them ideal for comparative analysis. The study's objectives include enhancing the efficiency of energy distribution systems, identifying peak demand periods, and supporting sustainable energy planning for large district areas. Additionally, the insights gathered can inform district-wide energy management practices and facilitate the integration of renewable energy sources, contributing to the development of smarter, more resilient urban energy systems.

This thesis develops and evaluates data-driven methodologies and advanced machine learning (ML) models for predicting, analyzing, and recommending energy load measures in urban district energy systems. It compares white box, black box, and hybrid modeling tools, with emphasis on their applicability in energy consulting practices. The research focuses on two major case studies: the University of British Columbia (UBC) campus and the city of Stockholm. For UBC, various ML and deep learning models, including decision trees, support vector regression, and artificial neural networks, were applied to forecast electrical, hot water, and gas consumption. These models achieved high predictive accuracy (e.g., R^2 up to 0.94, low MAE), and their performance was validated using a structured train-test-validate split. For Stockholm, spatial clustering methods (K-means, agglomerative clustering) were used to map heat demand and allocate residual heat sources. The levelized cost of heat (LCOH) was calculated per cluster to support infrastructure planning. The results demonstrate the effectiveness of data-driven modeling for both temporal forecasting and spatial energy optimization. The proposed methodologies are generalizable and can inform scalable, low-carbon district energy strategies.

Keywords: District energy modelling, White box models, Black box models, Machine learning, Energy load prediction

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

LIST OF ABREVIATIONS

AI
Artificial Intelligence

ANN
Artificial Neural Network

ASHRAE
American Society of Heating, Refrigerating, and Air-Conditioning Engineers

BEMS
Building Energy Management Systems

BES
Building Energy Simulation Tools

BIM
Building Information Modelling

BMS
Building Management Systems

BPS
Building Performance Simulation Tools

DBN
Deep Belief Network

DDM
Data-Driven Modelling

DL
Deep Learning

DNN
Deep Neural Networks

DT
Decision Tree

EPC
Energy Performance Certificate

GBR
Gradient Boosting Regressor

GHG
Greenhouse Gas

GIS
Geographic Information Systems

HVAC
Heating, Ventilation, and Air Conditioning

IFC
Industry Foundation Class

IoT
Internet of Things

KNN
K-Neighbors Regression

ML
Machine Learning

MSE
Mean Squared Error

NRMSE
Normalized Root Mean Square Error

RF
Random Forest

Ridge
Ridge Regression

RMSE
Root Mean Squared Error

R^2
Coefficient of Determination

SCOP
Seasonal Coefficient of Performance

SVR
Support Vector Regression

SVM
Support Vector Machine

UBC
University of British Columbia

ULTDHC
Ultra-Low Temperature District Heating and Cooling

Lasso
Lasso Regression

LR
Linear Regression

LSTM
Long Short-Term Memory

MAE
Mean Absolute Error

KTH
Royal Institute of Technology, Stockholm

LIST OF SYMBOLS

COP	Coefficient of Performance
TWh/year	Terawatt-hour per year
kWh/m^2	Kilowatt-hour per square meter
RMSE	Root Mean Square Error
EUR/MWh	Euro per Megawatt-hour
SCOP	Seasonal Coefficient of Performance for various heat sources
GWh/year	Gigawatt-hour per year

INTRODUCTION

As cities worldwide face increasing pressure to reduce greenhouse gas emissions, district energy management has emerged as a vital tool in transitioning toward sustainable urban environments (Renewable Energy Policy Network for the 21st Century, REN21, 2022). The urgency to combat climate change has prompted urban planners and policymakers to explore innovative energy solutions, with district energy systems (DES) at the forefront of these efforts. District energy systems, which include networks for heating and cooling, facilitate the efficient use of energy resources by centralizing and optimizing energy production and distribution across urban areas. These systems serve a variety of buildings, from residential complexes to commercial facilities, by providing them with heating, cooling, and hot water through a shared infrastructure. This approach enhances energy efficiency and promotes using renewable energy sources, reducing reliance on fossil fuels.

However, the complexity of these systems presents significant challenges for operational efficiency, cost-effectiveness, and emissions reduction (Digitemie & Ekemezie, 2024). The management of district energy systems involves navigating variable demand patterns, diverse energy sources, and fluctuating environmental conditions (Schuster, 2020). For instance, energy demand can fluctuate drastically based on the time of day, seasonal changes, and occupancy levels of the buildings served. Moreover, integrating renewable energy sources such as solar, wind, and biomass introduces additional variability, as these sources are subject to weather and environmental factors. The ability to effectively manage and optimize these complex systems is essential for minimizing energy waste, reducing operational costs, and achieving ambitious climate goals (Schuster, 2020).

Different modelling approaches, such as white box and black box models, are highly influential in addressing the complexities of energy optimization and system management. White box models are rooted in fundamental physics and engineering principles, using detailed mathematical representations to simulate physical processes and interactions within a system.

These models rely on a deep understanding of system mechanics and require extensive domain knowledge, making them valuable for scenarios where accuracy, interpretability, and detailed insights are crucial. Examples of white box models in energy management include building thermal models and system performance simulations, where specific conditions and parameters can be accurately controlled and evaluated (Amara et al., 2022).

Conversely, black box models, often referred to as data-driven approaches, do not require explicit knowledge of the underlying physical processes. Instead, these models rely on data patterns, machine learning algorithms, and statistical techniques to learn relationships within the data. Black box models, like neural networks and regression trees, are often more flexible, scalable, and capable of handling complex, nonlinear interactions that may be too intricate or time-consuming to model explicitly (Amara et al., 2022).

Black box models are particularly effective for real-time optimization, predictive maintenance, and operational energy management because they can rapidly learn from large datasets and adapt to changing conditions without fully understanding the system's mechanics. This data-driven approach is advantageous in solving the challenges of modern energy systems, as it allows for a quick and adaptive response to dynamic and diverse data sources, enabling more efficient and timely decision-making (Testasecca et al., 2023).

In addressing these challenges, black box models have gained prominence over traditional white box models and web-based tools (Amara et al., 2022). White box models provide transparency and interpretability by clearly outlining the underlying mechanisms of the system; however, they often struggle to capture the complexity and non-linear relationships present in large datasets (Amara et al., 2022). For example, a white box model might be able to predict energy demand based on a simple linear regression of historical data. Still, it would likely fail to account for the interactions between multiple variables, such as weather patterns, occupancy levels, and operational strategies influencing energy consumption.

Conversely, black box models excel in learning intricate patterns without requiring a detailed understanding of the underlying processes (Arendt et al., 2018). These models leverage advanced machine learning techniques to analyze vast amounts of data from various sources, including real-time sensor readings, historical consumption data, and external environmental factors. Their ability to process this information enables them to make accurate predictions and dynamically optimize operations in response to changing conditions. This data-driven flexibility is particularly crucial in district energy systems, where demand can be unpredictable and influenced by numerous variables.

Moreover, black box models have the advantage of being adaptive, continuously improving their accuracy and performance over time as they are exposed to new data. This adaptability is vital for the evolving needs of urban energy management, as it allows for incorporating real-time data and feedback into decision-making processes. For instance, a machine learning model could refine its predictions for energy demand as it learns from fluctuations in historical data, improving the responsiveness and efficiency of the district energy system (Arendt et al., 2018).

Data-driven approaches and machine learning models are increasingly integral to addressing the challenges faced by district energy systems. By utilizing large datasets from smart sensors, building management systems, weather data, and historical energy consumption, data-driven models can provide accurate and timely insights to enhance energy management. For example, predictive models can anticipate demand fluctuations based on historical patterns, temperature forecasts, and occupancy rates. This capability enables district energy systems to preemptively adjust supply levels, ensuring that energy is available when and where needed, thereby reducing waste and optimizing resource utilization.

Additionally, clustering and classification algorithms, such as K-means and Agglomerative Clustering, can map source-load relationships in district energy systems. These algorithms identify optimal locations for new energy sources or for utilizing waste heat, allowing city planners to pinpoint areas with high demand and locate local energy sources, such as data centers or supermarkets, that produce residual heat. For instance, K-means clustering can segment urban areas into distinct groups based on energy consumption patterns, helping to identify neighborhoods that would benefit from adding new heating or cooling sources. This clustering-based source-load mapping enhances energy efficiency by minimizing the distance between heat generation points and end users, thereby reducing transmission losses and improving overall system performance (Radtke, 2022).

Machine learning also plays a crucial role in evaluating and optimizing system performance over time. For example, reinforcement learning algorithms can dynamically adjust control strategies based on real-time feedback. In a district energy system, this could involve optimizing the operation of boilers, chillers, and pumps to maintain comfort levels while minimizing energy consumption. Reinforcement learning enhances system efficiency under changing conditions by continually learning from system performance and adjusting strategies accordingly (Idowu et al., 2014).

Furthermore, machine learning models such as artificial neural networks (ANNs) and decision trees can be utilized to forecast energy loads, detect anomalies, and assess the impact of various operational strategies on emissions and costs. ANNs, for example, can learn complex relationships between inputs and outputs, making them well-suited for predicting energy consumption based on a multitude of influencing factors. Decision trees, on the other hand, provide a more interpretable framework for decision-making, allowing stakeholders to understand the rationale behind specific operational adjustments and their potential impact on system performance (Borioli et al., 2009).

Moreover, district energy systems benefit from machine learning in predictive maintenance. By analyzing operational data and patterns of equipment wear, machine learning models can forecast potential equipment failures and schedule maintenance before unexpected breakdowns disrupt the system. This proactive approach reduces downtime, minimizing emergency repairs and lost service availability costs. For instance, a predictive maintenance model might analyze historical maintenance records and real-time sensor data to identify early signs of wear in a chiller, prompting timely intervention before a costly failure occurs (Mbiydzenyuy et al., 2021).

Integrating data-driven and machine-learning models with district energy management systems enhances operational efficiency while supporting broader environmental goals. By reducing energy waste, improving resource allocation, and enabling more sustainable use of residual heat sources, these advanced technologies play a crucial role in achieving carbon reduction targets and fostering resilient, sustainable urban communities. The potential benefits go beyond just improving operations; effective district energy management can also enhance energy security, support local economies, and improve the quality of life for urban residents. As cities face the challenges of climate change, using district energy systems supported by data and machine learning becomes more critical. The successful integration of these technologies has the potential to transform urban energy landscapes, making them more efficient, sustainable, and resilient. However, realizing this potential requires ongoing collaboration among city planners, policymakers, energy providers, and technology developers (Rehmann et al., 2024).

Efforts should be directed toward enhancing the capabilities of district energy systems, investing in research and development of advanced machine learning techniques, and fostering public awareness and engagement. As communities understand the benefits of sustainable energy practices, support for district energy initiatives increases, driving further innovation and investment in this critical area (Mbiydzenyuy et al., 2021). Moreover, policy frameworks should incentivize the development and deployment of district energy systems, ensuring that cities are equipped with the tools and resources necessary to transition to sustainable energy practices. By prioritizing investments in smart grid technologies, energy storage solutions, and renewable energy generation, municipalities can create an environment conducive to the growth of district energy systems (Agupugo et al., 2022). The primary objective of this thesis is to leverage machine learning (ML) and deep learning (DL) models to improve the prediction, evaluation, and recommendation of energy load measures in district and urban areas. By developing sophisticated data-driven models, this research aims to deliver actionable insights for optimizing energy consumption, enhancing load forecasting accuracy, and devising effective strategies for managing energy demand in urban environments. The analysis conducted within this work seeks to address the complex challenges of urban energy needs by creating adaptable and scalable predictive models that contribute to sustainable, efficient energy systems. Deep Learning (DL) is a subset of Machine Learning (ML) that uses multi-layered neural networks to learn complex patterns from large datasets. Unlike traditional ML, which often requires manual feature selection, DL can automatically extract high-level features, making it suitable for modeling non-linear energy consumption behaviors in this study.

To achieve these objectives, the thesis first focuses on predicting energy loads accurately. Load forecasting is essential for effective urban energy management, as fluctuating demands can lead to inefficient overcapacity or insufficient supply. Traditional methods often struggle to capture the intricate and nonlinear dynamics influencing urban energy consumption, such as weather patterns, occupancy rates, economic activities, and infrastructure diversity. Therefore, this thesis employs advanced ML and DL algorithms, including time-series models, regression trees, and neural networks, to generate precise, granular predictions of energy demand. These predictions form the foundation for energy system planning, enabling stakeholders to anticipate periods of high or low demand and facilitating improved resource allocation and infrastructure development (Mallala et al., 2024).

Another critical aspect of this research is evaluating model performance and reliability. The thesis aims to generate accurate predictions while rigorously assessing each model's effectiveness through various performance metrics, such as R-squared, Mean Absolute Error (MAE), and Root Mean Square Error (RMSE). This evaluation is crucial for identifying the models best suited to handle real-world complexities, such as seasonal variations and unexpected demand shifts, and for understanding each model's limitations. By comparing different model architectures, this research seeks to achieve an optimal balance between accuracy, interpretability, and computational efficiency, ultimately establishing a reliable framework for energy forecasting in urban contexts. In addition, this thesis advances data-driven solutions for urban energy by offering scalability, adaptability, and actionable insights. Urban energy systems are inherently complex, and their dynamic, interconnected components require a comprehensive approach to management. This research contributes to the growing body of work on data-driven solutions by developing models capable of processing diverse inputs, including weather conditions and building usage data, and demonstrating the transformative role of ML and DL in urban energy management. The overarching goal is to support cities' transition toward sustainable energy practices, aligning with the broader vision of smart cities that integrate digital technologies for an improved urban living experience.

This thesis's original contribution lies in applying advanced machine learning (ML) and deep learning (DL) methodologies to enhance the accuracy and reliability of energy load forecasting, evaluation, and optimization within district and urban areas. This work contributes a unique data-driven framework for urban energy management, addressing challenges such as non-linear consumption patterns, dynamic demand fluctuations, and the influence of external variables like weather and occupancy trends. By systematically evaluating various ML and DL models, including neural networks and regression trees, and refining them to capture urban energy complexities, this thesis provides novel insights into the adaptability and scalability of these methods within large-scale urban contexts.

Another key contribution is developing a rigorous model evaluation process that balances predictive accuracy with interpretability, establishing benchmarks for effective urban energy forecasting. Furthermore, this thesis expands upon traditional forecasting approaches by integrating diverse data sources and proposing targeted urban planning and energy distribution recommendations. This framework supports immediate energy load optimization and offers a replicable model that districts and cities can adopt to pursue sustainable energy systems, thus contributing to the growing research on smart cities and sustainable urban infrastructure. The insights generated from this work serve as a valuable resource for stakeholders in energy management, urban planning, and sustainability, bridging the gap between advanced predictive analytics and practical energy solutions for urban environments.

In conclusion, urban energy management is complex, but district energy systems offer significant potential for a sustainable future. By leveraging data-driven and machine-learning tools, cities can better manage energy production and distribution, lower greenhouse gas emissions, and improve residents' quality of life. The journey toward sustainable cities has challenges, but with the right tools, partnerships, and dedication, cities can pave the way to a greener, more sustainable future for everyone.

Chapter 1 details the methodology employed throughout the research, including the research framework, data collection strategies, and modelling approaches. It discusses the validation processes necessary to ensure the accuracy and reliability of the models and the application of the Elbow Method for optimal clustering in energy data analysis. The chapter concludes with an overview of model testing and evaluation metrics, providing a comprehensive understanding of the analytical framework that supports the subsequent findings.

Chapter 2 explores the transition from White to Black-Box Models in building energy management, reviewing various simulation tools and their applications in energy management systems. This chapter includes a classification of tools, an analysis of their integration into consulting, and assessments of their scalability. Examining both standalone and web-based models offers insights into the capabilities and limitations of existing simulation tools, ultimately providing recommendations for future research in this evolving field.

Following this, Chapter 3 focuses on advanced machine learning techniques used to analyze energy consumption and optimize performance at the UBC campus, correlating energy metrics with meteorological variables.

Chapter 4 discusses applying clustering techniques, specifically K-means and agglomerative clustering, for source-load mapping in district heating planning, highlighting the importance of efficient heat demand allocation.

The conclusion summarizes the key findings of the research and offers insights for future investigations into district energy systems and their management. This study highlights the advantages of machine learning and deep learning algorithms in optimizing energy forecasting and management. While white box models provide detailed, interpretable insights, they are computationally intensive and less scalable. In contrast, black box models enable rapid, accurate predictions, making them more suitable for real-time applications. The analysis of energy consumption at the UBC campus in Vancouver reveals strong correlations between energy sources and environmental factors, emphasizing the value of integrated energy management strategies. Advanced machine learning models, particularly deep learning, excel in capturing complex consumption patterns, supporting efforts to enhance efficiency and sustainability.

The research also explores Stockholm's district heating network, identifying optimal locations for integrating residual heat sources like data centers and supermarkets. Spatial and temporal data analysis suggests that waste heat could meet a significant portion of the city's heating demand, providing a cost-effective and sustainable solution. The study presents a scalable, data-driven framework for energy forecasting, balancing predictive accuracy and computational efficiency. These findings offer a replicable methodology for cities seeking to improve energy resilience and sustainability. Future research should further examine the causal relationships between energy consumption and external factors, assess operational interventions, and explore renewable energy integration to enhance urban energy planning.

Author Contribution

This thesis is based on a collection of publications. As the primary author of each included paper, the PhD candidate, Amir Shahcheraghian, made the following contributions:

1. From White to Black Box Models: A Review of Simulation Tools for Building Energy Management and Their Application in Consulting Practices (Energies, 2024)

- Conceived the study concept and structured the review methodology

- Conducted the literature review and comparative tool analysis
- Wrote the full manuscript and created all figures and tables
- Integrated consulting practice relevance and responded to peer reviews

2. Advanced Machine Learning Techniques for Energy Consumption Analysis and Optimization at UBC Campus: Correlations with Meteorological Variables (Energies, 2024)

- Led data preprocessing, feature engineering, and ML model development
- Designed and ran the experiments (regression, ANN models)
- Analyzed results, produced evaluation metrics, and wrote the manuscript
- Developed all figures/tables and handled revisions post-peer review

3. K-means and Agglomerative Clustering for Source-Load Mapping in Distributed District Heating Planning (Energy Conversion and Management: X, 2025)

- Formulated the clustering methodology and coordinated spatial data processing
- Implemented K-means and hierarchical clustering models using Stockholm EPC datasets
- Performed levelized cost of heat (LCOH) calculations and GIS visualizations
- Drafted and revised the manuscript and ensured methodological consistency across sections

All co-authors contributed through supervision, technical feedback, and critical manuscript review.

CHAPTER 1

METHODOLOGY

1.1 Research Methodology

This study aims to offer efficient methods for predicting, evaluating, and recommending energy load measures for use in district areas. By leveraging data-driven approaches, machine learning techniques, and artificial neural networks (ANNs), this research seeks to uncover complex relationships between energy consumption patterns and external variables such as weather, time of day, and building-specific factors. Integrating advanced analytics and predictive modelling enables more accurate energy demand forecasts, optimizing resource allocation and informing sustainable energy management strategies. These methods are valuable for decision-makers in urban energy planning, contributing to more efficient and environmentally responsible energy systems in district areas.

In the first step, acquiring knowledge about the various tools used in urban and district energy prediction is necessary. These tools can be categorized into three main types: white box models, which are based on fundamental principles and provide transparent insights into the system's workings; black box models, which utilize complex algorithms to make predictions without disclosing the underlying processes; and grey-box models, which combine elements of both white and black box approaches, integrating empirical data with theoretical foundations. Understanding the strengths and limitations of each model type is crucial for selecting the most appropriate approach for energy load prediction. In this study, black box data-driven models were chosen for their ability to capture intricate patterns and relationships within large datasets, thereby enhancing the accuracy of energy demand forecasts in urban and district contexts.

Ultimately, black box data-driven methods, such as machine learning (ML) and artificial neural networks (ANN), were chosen for this study due to their vast capabilities and exceptional proficiency in handling complex, nonlinear relationships within large datasets. These models excel at identifying hidden patterns and interactions that traditional modelling approaches may overlook, making them particularly suited for predicting energy load measures in urban and district contexts.

Their adaptability allows for continuous learning and improvement as more data becomes available, enhancing their predictive accuracy over time. Furthermore, the ability of ML and ANN to incorporate diverse data sources such as historical energy usage, meteorological variables, and building characteristics enables a comprehensive analysis of energy consumption dynamics. The study aims to provide actionable insights and robust recommendations for effective energy management strategies by leveraging these advanced methodologies. This study employed a systematic approach utilizing five main steps to effectively develop and implement the proposed energy prediction models:

- Data collection
- Modelling
- Validation
- Model testing and evaluation
- Optimization

1.2 Data Collection

An extensive investigation was conducted into various sources for the data collection phase to obtain suitable, reliable, and comprehensive datasets essential for the study. Notably, data was sought from Hydro Quebec in Quebec, Canada, but it was determined that the provided data did not meet the required quality and feature specifications. Similarly, data was requested from the Westmount municipality in Montreal and the Toronto municipality, but the necessary datasets were unavailable.

Open data from the U.S. Energy Information Administration (EIA) and the U.S. Department of Energy (DOE) was also examined; however, it was found that these sources did not possess the baseline quality and features needed for the study.

In contrast, comprehensive real-time data dating back to 2008 was provided by the UBC campus in Vancouver, encompassing energy metrics from 154 buildings on campus. This dataset was crucial for modelling efforts, as it offered granular insights into energy usage patterns under varying operational conditions.

Additionally, energy performance certificates (EPC) and coordinate data for Stockholm were contributed by KTH University, which included extensive information on energy consumption across the city. This data was invaluable for analyses and enabled the application of various machine learning (ML) and artificial neural network (ANN) models to identify relationships and patterns within the datasets. By leveraging these diverse and rich data sources, robust predictive models are aimed to be developed that effectively inform energy management strategies in urban and district areas.

To ensure data suitability for machine learning-based modeling and urban energy analysis, specific criteria were established for selecting datasets. First, the datasets needed to offer adequate temporal resolution, with hourly or daily measurements preferred, to support dynamic energy demand forecasting and time-series analysis. Second, the completeness of features was essential; only datasets that included core variables such as energy consumption (electricity, gas, or thermal), building characteristics (e.g., floor area, function), and, where applicable, environmental data (e.g., temperature, humidity) were considered. Third, data continuity and duration played a critical role; datasets were required to have long-term coverage (ideally more than one year) with minimal missing values to ensure reliable model training, testing, and validation. Finally, for spatial analysis tasks, particularly in the Stockholm case, geospatial granularity was necessary, including coordinate-based or district-level resolution to enable clustering and mapping. Datasets that failed to meet these criteria, such as those with aggregated monthly data, incomplete records, or insufficient spatial resolution, were excluded from the study. The datasets from the UBC campus and KTH University fully satisfied these criteria, offering the temporal, spatial, and variable detail necessary for robust machine learning and energy analysis across urban and district contexts.

1.3 Modelling

In the modelling phase, various machine learning (ML) models and artificial neural networks (ANN) were employed to predict, evaluate, and recommend energy load measures for district and urban areas. ML and ANN are particularly beneficial in energy consumption prediction due to their ability to capture complex, nonlinear relationships within large datasets, enabling a more nuanced understanding of how various factors impact energy demand. Several models were utilized in this study, including Decision Trees, which provide intuitive and interpretable results by splitting data into subsets based on feature values; Random Forests, an ensemble method that improves predictive accuracy by combining multiple decision trees to mitigate overfitting; and Gradient Boosting, which builds models sequentially, optimizing performance by focusing on errors made by previous iterations. Additionally, AdaBoost enhances the performance of weak classifiers by assigning weights to instances, allowing the model to focus on misclassified data points, thus improving accuracy.

For regression tasks, we applied Linear, Ridge, and Lasso Regression to analyze the relationship between features and energy consumption, where Linear captures basic trends and Ridge/Lasso adds regularization to manage overfitting. Support Vector Regression (SVR) provided robust predictions in high-dimensional, noisy data, while K-Neighbors used local averages to capture neighborhood effects. ANN models further captured complex patterns from historical data, factoring in influences such as weather, occupancy, and operations.

Collectively, these approaches form a comprehensive framework for urban energy load prediction, enhancing accuracy and delivering actionable insights for efficient, sustainable energy management.

1.4 Validation

In the validation phase, the performance of each model was rigorously evaluated using a structured data partitioning approach. For every model developed in this study, the dataset was divided into three distinct subsets: 70% for training, 15% for validation, and 15% for testing. This strategy ensured that models were trained on a sufficiently large portion of the data, while validation was used to tune hyperparameters and prevent overfitting where a model performs well on training data but fails to generalize to unseen data. The validation set guided model selection and performance monitoring during training, whereas the test set was reserved for final evaluation on completely unseen data. This three-way split improves the robustness of model assessment. Key evaluation metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R-squared (R^2) were employed to measure both prediction accuracy and model generalization. These metrics quantified how well each model captured the underlying energy consumption patterns and its effectiveness in forecasting future energy loads.

For models like Decision Trees and Random Forests, the validation process ensured that the models did not overfit the training data, which could lead to poor performance when generalized to new datasets. In ensemble methods such as Gradient Boosting and AdaBoost, the validation step was crucial to ensure that the iterative process of correcting errors did not lead to excessive focus on outliers. Similarly, for regression models like Linear Regression, Ridge Regression, and Lasso Regression, validation helped to identify whether regularization parameters needed tuning to balance bias and variance effectively.

In the case of artificial neural networks (ANN), the validation set was used during the model training process to monitor learning progress and adjust hyperparameters, such as learning rate and the number of hidden layers, to prevent the network from overfitting or underfitting. Early stopping techniques were also employed, where training was halted if the performance on the validation set stopped improving after a certain number of epochs.

The cross-validation technique was also considered, where the dataset was split into multiple folds, and each fold was used as a validation set once while the model was trained on the remaining folds. This method provided a more robust estimate of model performance by reducing the variance introduced by a single train validation split. Through the validation process, we ensured that the models developed in this study were accurate and capable of generalizing well to new data, making them reliable tools for predicting, evaluating, and recommending energy load measures in district areas.

In the validation process for clustering methods such as K-Means and Agglomerative Clustering, the approach differs from traditional supervised learning methods due to the unsupervised nature of clustering. The primary challenge in unsupervised learning is evaluating how well the clusters formed by the algorithm represent meaningful groupings in the data, as there are no predefined labels for comparison. To ensure the robustness and quality of the clustering results in this study, we employed a combination of internal and external validation techniques alongside a thorough analysis of cluster characteristics.

For clustering models like K-Means and Agglomerative Clustering, internal validation metrics help evaluate the quality of the clusters based on their compactness and separation. Two key internal validation metrics used in this study were the Silhouette Score and Inertia (within-cluster sum of squares). The Silhouette Score measures how similar a data point is to its cluster compared to other clusters, with values ranging from -1 to 1. Higher values indicate better clustering, where a score close to 1 signifies well-separated clusters, while scores near 0 suggest overlapping clusters. In this study, we utilized the silhouette score to assess the cohesiveness and separation of the clusters created by both K-Means and Agglomerative

Clustering methods. Conversely, Inertia represents the sum of squared distances between each data point and its nearest cluster center in K-Means, with lower values indicating more compact clusters. However, inertia alone does not determine cluster separation, so it is combined with other metrics like the silhouette score for a comprehensive evaluation. Additionally, dendrogram analysis was employed for Agglomerative Clustering (Thinsungnoen et al., 2015) to visually assess the hierarchy of clusters and select the optimal number of clusters. Dendrograms help us understand how clusters are formed through the iterative merging of smaller clusters, allowing us to determine a suitable threshold for stopping the clustering process and ensuring meaningful groupings.

Internal validation methods assessed cluster structure, while external validation metrics were used when labeled data was available. The Adjusted Rand Index (ARI) measured the similarity between generated clusters and ground truth, with values from -1 to 1, where 1 indicates perfect clustering. The Calinski-Harabasz Index evaluated cluster compactness and separation, aiding model comparison and cluster selection. For the district heating dataset, these metrics ensured clusters captured meaningful energy consumption patterns.

Through the validation process, we could fine-tune the parameters of both K-Means and Agglomerative Clustering to enhance the quality of the clusters. For K-Means, we adjusted the number of clusters based on the Elbow Method and Silhouette Scores to identify the optimal configuration for modelling energy consumption patterns. Similarly, in Agglomerative Clustering, we determined the appropriate number of clusters based on dendrogram cutoffs and the Calinski-Harabasz Index (Patel et al., 2022).

In conclusion, validating the clustering models was a multi-faceted process that incorporated quantitative evaluation metrics and qualitative analysis of the results in urban energy consumption patterns. By employing a combination of internal and external validation methods, we ensured that the models' clusters were statistically robust and practically meaningful for energy load prediction and optimization in district areas.

1.5 Clustering Selection and Validation

The Elbow Method is a widely used technique to determine the optimal number of clusters in clustering algorithms, particularly in K-Means clustering. This method helps strike a balance between underfitting and overfitting by analyzing the variance explained by different numbers of clusters. The first step involves running the K-Means algorithm on the dataset for a range of cluster values, such as from 1 to 10. For each value of k, the algorithm assigns data points to the nearest cluster centers and computes the within-cluster sum of squares (WCSS), also known as inertia. WCSS quantifies how tightly the clusters are packed together; lower WCSS values indicate more compact clusters.

After calculating the WCSS for each cluster count, these values are plotted on a graph, with the number of clusters k on the x-axis and the corresponding WCSS on the y-axis. This graphical representation allows us to visualize the relationship between the number of clusters and the compactness of the clusters. As k increases, the WCSS decreases, but a point is reached where the rate of decrease sharply declines, creating an "elbow" shape in the graph. This elbow point signifies that adding more clusters beyond this point results in diminishing returns in terms of explained variance, representing the optimal number of clusters where a good trade-off between complexity and accuracy is achieved.

This study applied the Elbow Method to the energy consumption dataset to ascertain the most suitable number of clusters for predicting energy load measures in district areas. Upon performing the clustering analysis and plotting the WCSS values against various cluster counts, it was observed that the elbow point occurred at k=4. Several factors justified this choice of 4 clusters. First, the analysis indicated that with 4 clusters, the WCSS significantly dropped, suggesting that the data points were well-separated and tightly grouped within each cluster, which is crucial for effective clustering in energy load predictions. This compactness allows for better interpretation and analysis of energy consumption patterns.

Moreover, using 4 clusters provided a manageable and interpretable model that aligns well with practical considerations in energy management. Each cluster can represent a distinct energy consumption category, facilitating the development of targeted energy efficiency and load optimization strategies. Selecting more clusters, such as 5 or more, could lead to a more complex model with potential overfitting, capturing noise rather than actual patterns in the data. The elbow method confirmed that 4 clusters balanced the need for detail in analysis while maintaining model simplicity.

In addition to using quantitative metrics, this study's validation of clustering results also relied heavily on domain expertise. Specifically, for the district heating and energy consumption data, the clusters were interpreted in the context of real-world geographic locations and energy consumption patterns. For example, in the Stockholm case study, we assessed whether the clusters identified by the models aligned with known areas of high heat demand, potential heat recovery sources, or zones suitable for new data centers.

By integrating domain knowledge with quantitative validation metrics, we ensured that the clusters exhibited statistical validity and practical relevance for energy planning and district heating optimization.

1.6 Model Testing and Evaluation

The methodology involves extensive testing and evaluation of various forecasting models to assess their ability to predict energy consumption. This process includes applying different regression techniques, such as decision tree regression, to test the models across multiple years, allowing for a comprehensive evaluation of predictive accuracy over various time frames.
To measure the effectiveness and reliability of each model, several performance metrics are employed:

R-squared (R^2): R^2 measures the proportion of variance in energy consumption explained by the model's features. Ranging from 0 to 1, higher values indicate a better fit. It should be interpreted with other metrics to avoid overestimating performance. Similar studies have used R^2 to evaluate how well the model captures variability in energy consumption. The formula for R^2 is:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{1.1}$$

y_i is the actual value of the dependent variable.

$\hat{y}_i$ is the predicted value of the dependent variable.

$\hat{y}_i$ is the mean of the actual values of the dependent variable.

n is the number of observations.

Mean Absolute Error (MAE): MAE quantifies the average error magnitude between actual and predicted values. Lower MAE indicates higher accuracy, making it a widely used metric for model evaluation. However, MAE does not penalize large errors more than smaller ones. The formula for MAE is:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{1.2}$$

Mean Squared Error (MSE): MSE quantifies the average squared difference between the predicted and actual values. It is one of the most widely used loss functions in regression tasks due to its sensitivity to large errors, making it particularly useful for penalizing large deviations. Lower MSE values indicate better predictive accuracy. In energy modeling studies, including those focused on building energy demand forecasting, MSE helps assess how closely model predictions align with actual energy consumption patterns.

$$MSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{1.3}$$

By analyzing these metrics, the methodology assesses each model's performance in capturing energy consumption patterns, providing valuable insights into their predictive accuracy.

CHAPTER 2

From White to Black Box Models: A Review of Simulation Tools for Building Energy Management and Their Application in Consulting Practices

Amir Shahcheraghian [a], Hatef Madani [b], Adrian Ilinca [a]

[a] Department of Mechanical Engineering, École de Technologie Supérieure, 1100 Notre-Dame West, Montreal, Quebec, Canada H3C 1K3
[b] Department of Energy Technology, KTH Royal Institute of Technology, SE-100 44 Stockholm, Sweden

Paper Published in *MDPI-Energies*, January 2024

Abstract
Buildings consume significant energy worldwide and account for a substantial proportion of greenhouse gas emissions. Therefore, building energy management has become critical with the increasing demand for sustainable buildings and energy-efficient systems. Simulation tools have become crucial in assessing the effectiveness of buildings and their energy systems, and they are widely used in building energy management. These simulation tools can be categorized into white box and black box models based on the level of detail and transparency of the model's inputs and outputs. This review publication comprehensively analyzes the white box, black box, and web tool models for building energy simulation tools. We also examine the different simulation scales, ranging from single-family homes to districts and cities, and the various modelling approaches, such as steady-state, quasi-steady-state, and dynamic. This review aims to pinpoint the advantages and drawbacks of various simulation tools, offering guidance for upcoming research in the field of building energy management. We aim to help researchers, building designers and engineers better understand the available simulation tools and make informed decisions when selecting and using them.

Keywords: BES; simulation tool; white box ; black box ; machine learning; deep learning; building energy

2.1 Introduction and Motivation

The field of building energy management is undergoing a transformative evolution, driven by the ever-increasing need for sustainable and energy-efficient solutions in the construction and operation of buildings. Recent studies like those by Doe and Smith (Doe, 2023) illuminate the potential of cutting-edge simulation software, underscoring a significant shift towards advanced tools that enable real-time optimization of energy use.

As energy efficiency and sustainability become paramount in building design, operation, and retrofitting, the demand for accurate, data-driven decision-making has never been higher.
In this dynamic landscape, simulation tools have emerged as indispensable, offering professionals the means to model, analyze, and optimize the energy performance of buildings. This review publication explores the diverse spectrum of simulation tools available for building energy management, highlighting their strengths, limitations, and applicability, focusing on their integration into consulting practices. Zhao and Patel's (Zhao, 2022) work on incorporating machine learning into building energy models exemplifies the progression toward data-driven methodologies revolutionizing energy management practices.

The scope of this review encompasses two fundamental categories of simulation tools: white box models and black box models. White box or physics-based models rely on fundamental physics principles and engineering equations to simulate the intricate interactions within a building's energy systems. In contrast, black box models leverage empirical, data-driven approaches, often incorporating machine learning algorithms and statistical techniques to predict building energy consumption and behaviour.

Consulting practices in building energy management necessitate a deep understanding of these simulation tools and their alignment with specific project requirements. As documented by Huang and Nguyen(Huang, 2023), the practical application of these tools in consulting practices is theoretical and a reality that is reshaping the industry, revealing a paradigm shift in building energy management. Their work underscores the critical role of these tools in delivering actionable insights and the added value they bring to energy consulting firms.

This review delves into the technical intricacies of these tools, highlighting their applications in consulting scenarios. Whether it is conducting energy audits, optimizing building systems, or ensuring compliance with energy performance standards, these tools have become indispensable assets for consultants. By presenting case studies and industry insights, we showcase the tangible impact of simulation tools on building energy management, contributing to the continuous improvement of energy efficiency and sustainability in the built environment.

Ultimately, this review presents a contemporary survey of white box and black box simulation tools in building energy management, focusing on their consulting applications. Our detailed comparative analysis equips professionals and researchers with insights for informed decision-making, energy efficiency optimization, and fostering a sustainable built environment.

2.2 Literature Review

Statistically, cities are among the largest energy consumers and greenhouse gas emitters (Poponi et al., 2016). Therefore, predicting building energy is vital for strategizing and enhancing energy systems (Huang, 2023; Zhao, 2022) and the penetration of renewable energy (Ahmad & Chen, 2019; Salkuti, 2019). It is crucial to lower energy usage in buildings, boost efficiency, and raise the proportion of renewable energy consumption.

As energy becomes increasingly critical to countries' economies and the environment, considerable efforts are made worldwide toward its optimal use and sustainable development. The problem is associated with an energy "trilemma," defined as the need to improve the security of supply, human comfort, and accessibility. The energy is in a complex interaction with other resources like water and land. Competing demands require reducing energy costs to consumers and reducing carbon emissions for a minimal increase in the global average surface temperature (Harris, 2016). Also, the load and energy management systems directly affect the occupant experience in commercial and residential buildings (Fan et al., 2017).

Due to the rapid growth of the city's inhabitants and the 40% share of building energy in energy consumption (Arendt et al., 2018), it is inevitable to improve building efficiencies. Energy consumption modelling is the first step to analyzing and optimizing building efficiency. Several building energy consumption modelling tools have been developed in the last twenty years, ranging from data-driven models to web tools. Sandra et al. evaluate the effectiveness, specifically in terms of accuracy and robustness, of 60 calibration methods based on optimization for white box models (Sandra et al., 2020). Zhengwei et al. assess approaches for comparing building energy use with its historical or expected performance, and they analyze the differences between white box and grey-box models (Li et al., 2014). Finally, Xiwang et al. examine recent advancements in building energy modelling, encompassing both comprehensive building and key component modelling, for building control and operation. They discuss and compare various methods, ranging from white box to black box models (Li & Wen, 2014).

Building energy tool selection criteria depend on factors like inputs and outputs, building or district analysis, etc. An analysis of the building performance using a new evaluation method is presented (Migilinskas et al., 2016). This article determines the impact of intricate factors such as construction duration, construction expenses, annual costs based on bills, primary energy requirements, yearly CO_2 emissions from energy usage, CO_2 emissions from construction materials and activities, and thermal comfort on ultimate decision-making.

Occupant behaviour is the next factor that can affect tool selection. Delzendeh et al.'s review seeks to determine prevailing research directions, pinpoint unexplored areas for future study, and identify trends in prominent journals using the Science Direct and Scopus databases (Delzendeh et al., 2017).

Eva Schito et al. explore methodologies and technologies to reduce building energy requirements, emphasizing the significant potential for savings through retrofits amid global efforts to lower consumption and carbon emissions. They discuss the impact of EU directives, building design, renewable integration, and multi-objective optimization in sustainable solutions, highlighting research that balances energy efficiency with economic, architectural, technological, and human comfort factors (Schito & Lucchi, 2023).

Gwanggil Jeon discusses the increasing role of AI models in energy management and decision-making. Various AI applications in energy systems, such as renewable energy estimation, demand forecasting, and optimization of energy consumption in public transportation, are covered. Enhanced efficiency, accuracy, and predictive capabilities are achieved through AI use in these areas, offering robust solutions for energy-related challenges. Contributions integrating AI with existing energy systems are featured in the document, showcasing AI's potential to bring stability, security, and efficiency to the energy sector (Jeon, 2022).

Yiqun Pan et al. aimed to identify and organize the appropriate principles, methods, and tools for engineers and researchers involved in building energy management, together with case studies that could hold academic or practical importance (Pan et al., 2023a). Therefore, the review was organized into five sections, each aligning with distinct goals of building performance simulation. These sections include performance-driven design, operational performance optimization through modelling, integrated simulation with data measurements for digital twin creation, building simulation aiding urban energy planning, and modelling building-to-grid interactions for the demand response (Yiqun Pan et al., 2023a).

Abdo Abdullah Ahmed Gassar et al. offer a comprehensive summary of past research efforts to forecast large-scale building energy consumption through diverse methodologies, encompassing black box, white box, and grey-box techniques. This review covers various facets of large-scale building energy prediction, including elements influencing building energy requirements, different building categories like residential, commercial, and office structures, and prediction ranges extending from a cluster of buildings to an entire city, region, or nation(Gassar & Cha, 2020).

The exploration of energy efficiency, renewable energy utilization, and environmental protection by Francesco Calise et al. is presented. Research from the International Conference on Sustainable Energy and Environmental Development (SEED) is showcased, including hybrid renewable energy systems, organic Rankine cycle enhancements, solar collector performance, and microgrid system design. The importance of integrating technological, economic, and environmental perspectives to meet the challenges of sustainable energy development and environmental protection is emphasized in this research (Calise & Figaj, 2022). Mohamed-Ali Hamdaoui et al. review two models for simulating hygrothermal behaviour in hygroscopic material buildings: white box and black box models. White box models, utilizing software like COMSOL Multiphysics (V5.6) or WUFI (V6.7), focus on physical understanding and balance equations. In contrast, black box models rely on statistical methods (ANN, CNN, LSTM) using measured data. The paper categorizes white box models into the CFD approach, with multiple control volumes per zone, and the nodal approach, treating each zone as a uniform volume (Hamdaoui et al., 2021). Xiaoliang Zhang et al. investigate the applications of the building simulation tool DeST (Design Simulation Toolkit) in building design and building energy efficiency research and consultation. They highlight how DeST has been used in various projects, including the development of building regulations and scientific research. The paper details DeST's role in building design consultation, commissioning, energy conservation assessment, and a building energy labelling system.

They present examples from a demonstration building to illustrate how DeST aids in design processes. Additionally, the paper mentions its use in other projects and regulations, demonstrating the widespread application of DeST in building energy efficiency (Zhang et al., 2008).

While the literature provides a comprehensive overview of various simulation tools, from the physics-based white box models to the data-driven black box approaches, there remains a notable disconnect between the theoretical capabilities of these tools and their practical application in consulting practices. Studies emphasize technical specifications and theoretical improvements in energy modelling. Still, they must address the real-world challenges consultants face, such as tool interoperability, user-friendly interfaces, and actionable outputs for decision-making. Moreover, there needs to be more comparative analysis that critically evaluates the performance of these tools in live consulting environments across different building types and energy systems. Furthermore, while advancements in areas like artificial intelligence present new opportunities for tool enhancement, their potential impact on consulting practices still needs to be explored. Future research must bridge these gaps by focusing on the usability of simulation tools in consulting practices, developing case studies that demonstrate their effectiveness in diverse scenarios, and assessing how emerging technologies can be harmoniously integrated to advance both the state-of-the-art and the practicalities of energy management consulting.

2.3 Objectives and Methodology

Efficient management of energy resources in the built environment is critical to sustainable urban development and environmental conservation. As the demand for energy-efficient buildings continues to grow, the role of simulation tools in building energy management becomes increasingly significant. These tools enable professionals to model, analyze, and optimize the energy performance of buildings, ultimately leading to reduced energy consumption and environmental impact. The review centers on these principal elements to realize the goals outlined in the introduction.

2.3.1 Tool Classification

This review systematically categorizes simulation tools for building energy management into two primary classes: white and black box models. These classifications serve as a foundational framework for comprehending the diverse modelling methodologies employed by these tools. White box models, often called physics-based models, are rooted in fundamental physics and engineering principles. They meticulously simulate building energy performance by considering the physical behaviour of various components and systems within a building. Prominent examples of white box tools include EnergyPlus, TRNSYS, and IDA-ICE. In contrast, black box models operate on empirical, data-driven approaches, frequently integrating machine-learning algorithms and statistical techniques. These models harness historical data to predict building energy consumption and behaviour. Well-known black box tools encompass Support Vector Machines, Random Forest, and Deep Neural Networks. This classification forms a structured basis for our analysis and highlights the fundamental differences between these modelling approaches. White box models aim to deeply understand the physical processes governing energy consumption, while black box models prioritize prediction accuracy, even if they are less interpretable. This fundamental distinction is pivotal in selecting appropriate tools for specific building energy management tasks.

2.3.2 Cosulting Practices Integration Analysis

The second critical aspect of our review explores how these simulation tools are integrated into consulting practices within the realm of building energy management. Consulting in this context refers to the professional services offered to clients seeking energy-efficient solutions for their buildings, whether they are new construction projects or existing structures requiring retrofitting. Consulting practices in building energy management encompass various activities, including energy audits, predictive modelling, retrofit analysis, and performance verification, as shown in Figure 2.1 (Lee et al., 2015).

The integration of simulation tools into consulting practices presents both challenges and advantages. White box models, such as EnergyPlus and TRNSYS, offer a detailed understanding of building energy systems but demand extensive inputs and calibration efforts. Their application is particularly suited for projects where a thorough comprehension of energy behaviour is critical, such as retrofitting projects and complex building systems optimization.

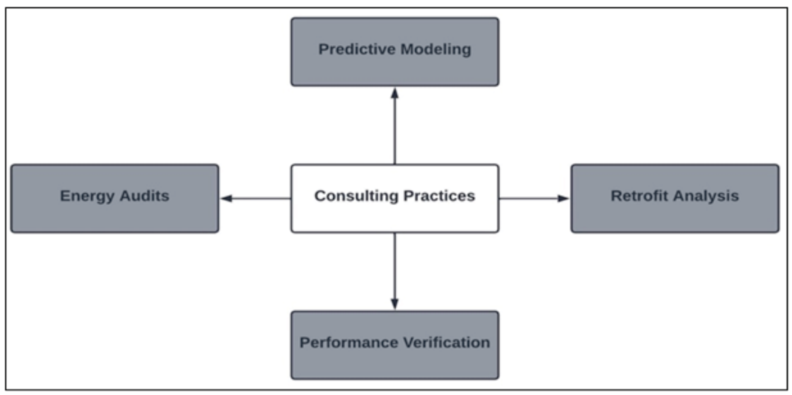

Figure 2.1 Specific features of consulting practices

On the other hand, Black box models like SVMs, Random Forests, and Deep Neural Networks deliver fast results with minimal input but require transparency and customization. They excel in predicting energy consumption, anomaly detection, and optimizing building operations, making them ideal for quick decision-making in consulting.

2.3.3 Scalability Assessment

The third aspect of our review assesses the scalability and adaptability of these tools. Building energy management projects can vary significantly in scale, ranging from individual building components to entire urban areas (Allegrini et al., 2015). Therefore, it is crucial to evaluate the suitability of each tool for different project sizes and complexities. White box models, with their in-depth simulations, are well-suited for complex building systems and projects where detailed modelling is essential.

They can accurately capture the interactions between various building components and systems, making them valuable for optimizing energy use while maintaining occupant comfort. With their data-driven approach, black box models offer flexibility and speed in consulting practices. They can be applied to various projects, from small-scale energy audits to large-scale urban energy analysis (Hopfe et al., 2017). Their ability to handle high-dimensional data and capture complex relationships between variables makes them adaptable to various building typologies and external influences, such as weather patterns. By systematically evaluating the scalability and adaptability of these tools, our review aims to assist professionals and consultants in selecting the most appropriate tool for their specific consulting projects. Whether the goal is to optimize the energy performance of a single building or develop sustainable urban energy strategies, choosing the right tool is essential for achieving accurate and actionable results.

2.3.4 Analytical Procedure

To achieve our research objectives, we have devised a systematic methodology that ensures the reliability and comprehensiveness of our review. Firstly, we conducted an extensive literature review encompassing peer-reviewed research articles, industry reports, and publications on building energy management simulation tools. This thorough examination guarantees that our analysis is firmly grounded in this domain's latest advancements and industry practices. Next, we categorized the identified simulation tools into two fundamental classes: white box and black box models. This categorization is based on these tools' underlying principles and methodologies, serving as a foundational framework that structures our analysis. It enables a clear understanding of the strengths and limitations of each tool category. Furthermore, our methodology includes an in-depth exploration of how these simulation tools are applied in the context of consulting practices. This involves meticulously examining case studies, industry insights, and best practices. Additionally, we conducted a scalability assessment of each tool, considering factors such as the level of detail they offer and their suitability for diverse consulting scenarios. This analysis aids in determining the applicability of these tools to a wide range of project scales.

Through the systematic execution of this methodology, our review aims to offer valuable insights into the evolving landscape of simulation tools for building energy management. We present a holistic perspective on these tools' capabilities, advantages, and limitations, empowering professionals in the field to make well-informed decisions and enhance the quality of their consulting services. The system boundary in building energy management tools refers to the level of analysis and detail the tool provides. As a result, building energy management tools can operate at different system boundaries, ranging from individual building components to entire urban regions (Figure 2.2).

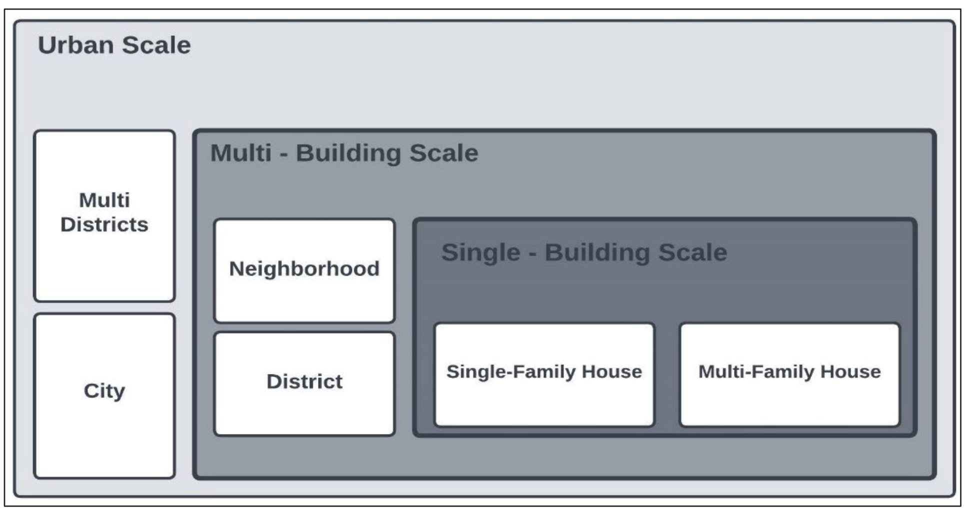

Figure 2.2 System boundary in BEM Tools

At the building system boundary, tools consider the interaction between various building systems and how they affect the overall energy performance (Allegrini et al., 2015). In addition, these tools often provide energy simulations and optimization capabilities to inform building design and operation decisions (Pan et al., 2023b). At the city or regional scale, tools consider large-scale urban systems' energy and environmental impacts, such as transportation, land use, and energy infrastructure. These tools may inform policy decisions about energy and climate change mitigation strategies (Sola et al., 2018). The choice of system boundary depends on the specific application and modelling objectives. BEM tools can be used at multiple system boundaries to inform design and operation decisions and assess the potential for energy savings and emissions reductions (Sola et al., 2018).

This review publication examines and compares standalone and web-based models in different resolutions, from white box to black box, in the context of building energy simulation tools (Figure 2.fi3). Building energy simulation is an essential instrument for analyzing and enhancing energy efficiency in buildings. White box models, grounded in fundamental physics and detailed component representations, thoroughly comprehend energy dynamics. Conversely, black box models rely on empirical data relationships for simulating energy consumption patterns. Furthermore, we underscore the importance of web-based models, emphasizing their inclusion within both white box and black box modeling paradigms. By exploring the strengths and limitations of these models, we seek to provide valuable insights and guidance to researchers, practitioners, and stakeholders involved in building energy management. The review will contribute to a better understanding of the diverse modelling options available and assist in selecting the most appropriate approach based on specific project requirements and constraints.

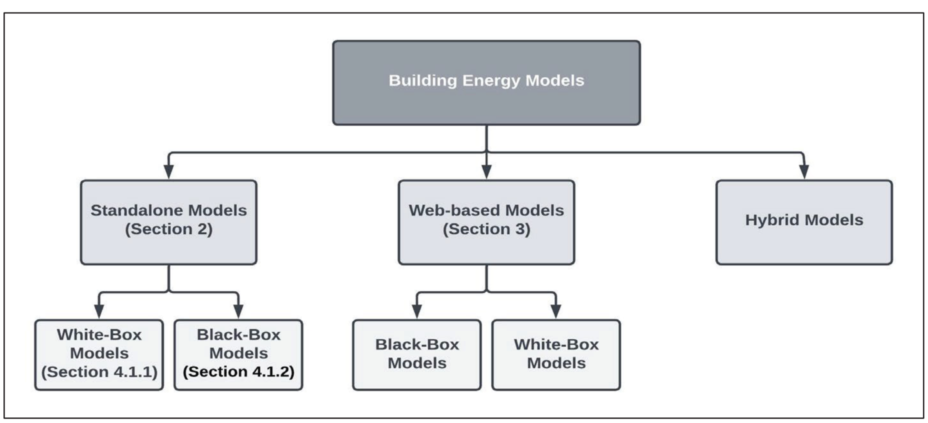

Figure 2.3 Structure of this review

2.4 Simulation tools for BEM

This section presents and categorizes simulation tools used in Building Energy Modeling (BEM), highlighting their underlying modeling principles, capabilities, and application domains. The tools are classified to support informed selection in both research and consulting contexts.

2.4.1 Standalone Models

Standalone models operate independently and often focus on detailed simulations of energy flows, thermal dynamics, and system interactions at the building or district level. These models are typically physics-based and require detailed input parameters for high-fidelity results.

2.4.1.1 White Bix Models

White box tools in building energy simulation refer to software programs that use explicit mathematical models of building components and systems to simulate energy consumption and other building performance metrics.

These models are derived from fundamental physical principles and engineering laws. In addition, the user can use white box modeling to dimension specific arrangements and test the data and parameters of the different scenarios (Hadziomerovic, 2019). A list of white box models can be found in Table 1.

White box tools can be more complex than black box tools and may require specialized knowledge and expertise to operate effectively. They may also require more detailed and accurate data inputs to produce accurate results, which can be challenging in some situations. They can predict energy consumption by establishing long-term associations between buildings' energy usage and important influencing factors (Ferrando et al., 2020). Energy models based on physical principles are the most accurate, and this approach is used by software such as DOE-2 and EnergyPlus. White box models, however, are complicated to build since they must include all the required equations and data. They need the most computing power and complexity, making their simulations slow.

Using white box techniques can alleviate many inefficiencies and resource-intensive characteristics of conventional complex models. Besides their accuracy, white box models offer the advantage of not requiring historical data. If the physical properties of a building are known, they can simulate a new building that does not exist (Arendt et al., 2018).

The white box building models apply heat and mass balance equations, which dynamically describe building behaviour. These equations account for three heat transfer mechanisms (conduction, convection, and radiation) between the building envelope and its surroundings. Numerous commercial and open-source software options, including EnergyPlus, Dymola, TRNSYS, and DOE-2, are available for building energy modelling. These tools efficiently formulate and solve these equations, although manual calculation of cooling and heating loads may still be necessary (Chen et al., 2022). The details for building thermal and cooling load prediction are presented in reference (A.Bhatia et al., 2015).

To create these models, comprehensive building data is essential, encompassing details such as building envelope characteristics, HVAC system configurations, internal heat contributions, equipment specifications, occupancy patterns, thermal zones, geographical location, and meteorological data, all of which are used to construct a physical building energy model (Drury B.Crawley, 2001). Predictive analysis and energy auditing are essential components in the consulting industry, especially regarding building efficiency (Akguc et al., 2013). White box tools play a crucial role in these processes. These tools enable consultants to conduct thorough energy audits and predictive analyses of buildings, modelling the energy behaviour of various components with high accuracy. This precise modelling is invaluable, particularly in retrofitting projects, where understanding the impact of modifications on energy performance is critical. White box tools offer a significant advantage in design and retrofitting decisions.

They allow consultants to simulate multiple design scenarios, providing clients with clear, data-driven insights regarding the energy implications of different design choices. This capability is crucial in the early stages of building design, where decisions can significantly influence future energy consumption. For retrofitting existing buildings, these tools are instrumental in evaluating the effectiveness of various energy-saving measures, such as upgrading insulation or HVAC systems (de Wilde & Augenbroe, 2018). White box tools are invaluable when optimizing building systems, such as HVAC, lighting, and ventilation. They can simulate the dynamic interactions between these systems and the building envelope, enabling consultants to devise strategies that boost overall energy efficiency while maintaining or improving occupant comfort.

Compliance and performance verification are other critical areas where white box tools are used. Many regions have specific energy performance standards for buildings, and these tools help consultants ensure that designs comply with these standards. They provide detailed analyses demonstrating compliance and are also used in performance verification to ensure that buildings operate as intended, achieving the designed energy efficiency levels. White box tools are particularly suited for modelling complex buildings, which may have unique architectural features or advanced energy systems. Their detailed nature allows consultants to develop customized solutions that address specific challenges, such as unusual building geometries or integrating renewable energy systems (de Wilde & Augenbroe, 2018).

Table 2.1 White-Box Models

Software	Version	Developer	City, Country	Platform	Timeframe	System Boundary	Available Outputs				Pros and Cons	References
							Energy	Thermal	Daylight	Air Quality		
EnergyPlus	23.2.0	DOE, NREL	Golden, CO, USA	Win, Mac, Linux	Sub-hourly, Hourly, user-defined timeframe	Neighborhood and Districts	✓	✓	✓	✓	Pros: Highly accurate for a variety of simulations, widely used and supported. Cons: Steep learning curve,	(Energy, 2015)
TRNSYS	18.03.0000	University of Wisconsin	*Madison, WI,* USA	Win	Dynamic (down to 0.01 s time-steps)	Neighborhood and Districts	✓	✓	✓		Pros: Flexible with a modular approach, good for both simple and complex systems. Cons: Requires in-	(TRNSYS)
City Sim	10 October 2023	EPFL Uni	Zurich, Switzerland	Win	Dynamic (hourly basis)	Multi-district and cities	✓				Pros: Specialized for urban-scale simulations, good for assessing microclimates and district energy	(E. University)
IDA ICE	5.0	EQUA Simulation	Glasgow, Scotland, UK	Win	Hourly	Neighborhood and Districts	✓	✓	✓	✓	Pros: Detailed thermal comfort and indoor climate simulations, user-friendly interface. Cons: License cost can be high, less	(AB; IBPSA-USA)
Envi-met	5.6	ENVI-met GmbH	Bochum, Germny	Win	Hourly	Single-Family and Multi-Family House	✓	✓			Pros: Strong for outdoor microclimate analysis and urban areas, good visualization tools. Cons: Focused more	(met)
Energy Pro	4.0361	EnergySoft	Novato, CA, USA	Win	Hourly	Multi-district and cities	✓	✓	✓		Pros: Certified for Title 24 compliance, user-friendly for architects and professionals. Cons: Primarily	(EnergySoft)
Retscreen	Version 9	Gov of Canada	Ottawa, ON, Canada	Win	Monthly basis (maximum: 50 years)		✓				Pros: Simplified tool for feasibility analysis and efficiency measures, includes climate data. Cons: Not as	(Canada)
EnerGis	8.1	EnerGis	-	Win	Monthly		✓	✓				(EnerGis)

Software	Version	Developer	City, Country	Platform	Timeframe	System Boundary	Energy	Thermal	Daylight	Air Quality	Pros and Cons	References
HOMER	3.10	UL	CO, USA	Win	Dynamic (minimum time-step 1 min)	Single-Family and Multi-Family House	✔	✔			Pros: Well-suited for optimizing microgrid designs, great for handling off-grid and renewable energy system simulations.	(Pro)
Neplan	10.940	NEPLAN AG	Zurich, Switzerland	Win	Hourly	Multi-district and cities	✔	✔				(AG)
Radiance	6.0a	Greg Ward	Berkeley, CA, USA	Win, Mac, Linux	Dynamic	Single-Family and Multi-Family House			✔		Pros: Highly accurate for daylighting and lighting simulation. Cons: Complex to use and requires significant	(Fritz)
Solene	microclimat	French National Research Agency (ANR) and	Lausanne, Switzerland	Win	Hourly	Single-Family and Multi-Family House	✔	✔	✔		Pros: Specialized in urban physics, it helps analyze solar radiation and its effects on buildings and urban spaces. Cons: May	(Imbert et al., 2018; Morille et al., 2015)
ESP-r	13.2.1	University of Strathclyde	Glasgow, Scotland, UK	Win, Mac	Hourly, Weekly, Monthly	Single-Family and Multi-Family House	✔	✔		✔	Pros: A versatile simulation environment capable of detailed thermal analysis, including HVAC and	(University of Strathclyde)
Be10	Specific version not available	DBRI	-	Win	Hourly	Single-Family and Multi-Family House	✔				Pros: Widely used in Denmark, particularly for compliance with Danish building regulations, user-friendly with a clear	(Institute, 2015)
BSim	Specific version not available	DBRI	-		Hourly	Single-Family and Multi-Family House	✔	✔	✔	✔	Pros: Comprehensive approach to simulating indoor environment and energy consumption in buildings. Cons: tailored	(Danish Building Research
DOE2	DOE-2.3 (release candidate)	James J. Hirsch & Associates (JJH)- eQuest	USA	Win	Hourly	Neighborhood and Districts	✔		✔		Pros: Provides a detailed and reliable simulation of building energy usage, with a strong emphasis on accuracy for HVAC and	(*EnerLogic and James J. Hirsch &*
IESVE	IESVE 2023	IES	Glasgow, Scotland, UK	Win	Hourly	Neighborhood and Districts	✔	✔	✔	✔	Pros: Comprehensive suite of tools for building performance simulation, strong for compliance and detailed HVAC	(IES)

Software	Version	Developer	City, Country	Platform	Timeframe	System Boundary	Energy	Thermal	Daylight	Air Quality	Pros and Cons	References
Velux	3.0	Velux Group	Horsholm, Denmark	Win, Mac	Hourly	Single-Family and Multi-Family House			✓		Pros: Renowned for daylighting capabilities and design. Cons: Focuses primarily on daylighting solutions and may not	(Velux)
iDbuild	-	Aarhus Uni	-	Win, Mac	Hourly	Single-Family and Multi-Family House	✓	✓	✓	✓	Pros: Offers an integrated approach to energy, indoor climate, and cost analyses. Cons: Can be complex due to its broad scope, which	(Petersen, 2013)
Design Builder	7.0	Design Builder Software Ltd.	Glasgow, Scotland, UK	Win	Hourly	Neighborhood and Districts	✓	✓	✓	✓	Pros: User-friendly interface, integrates simulation and building modelling with good visualization. Cons: May not	(Ltd)
eQuest	3.65	eQuest	USA	Win	Hourly, Weekly, Monthly	Neighborhood and Districts	✓				Pros: Free and widely used, particularly in the U.S. Cons: Interface can be less intuitive, and customization may be limited	(Associates)
OpenStudio	360	NREL	CO, USA	Win, Mac, Linux	Hourly, Weekly, Monthly	Multi-district and cities	✓	✓	✓	✓	Pros: Integrates with EnergyPlus and SketchUp, offering a more user-friendly interface for these powerful engines.	(Laboratory)
Riuska	4.9		-			Neighborhood and Districts	✓	✓		✓	Pros: Designed for climate analysis, offering detailed insights into microclimate and urban heat island effects.	(Granlund)
Sefaira	Sefaire 2018	Sefaira	London, UK	Win	Hourly, Weekly, Monthly	Multi-district and cities	✓	✓	✓		Pros: Known for its real-time energy and daylighting analysis within the early stages of design, providing	(Sefaira, 2015)
DIVA	4.0	Rhino	Cambridge, MA, USA	Win	Hourly	Single-Family and Multi-Family House			✓		Pros: Integrates with Rhino and Grasshopper, excellent for daylight and solar analysis with a visual programming interface.	(Solemma-LCC)
WatchWire	-	Energy Watch	-	Win, Mac	Hourly	Neighborhood and Districts	✓	✓		✓	Pros: Provides energy tracking and analytics geared towards operational energy management. Cons:	(WatchWire)
Sky Spark	3.1	SkyFoundry	Richmond, VA, USA	Win, Mac, Linux	Hourly	Multi-district and cities	✓				Pros: Excellent for data analytics and monitoring. Cons: More focused on data after buildings are operational.	(SkySpark)

Software	Version	Developer	City, Country	Platform	Timeframe	System Boundary	Energy	Thermal	Daylight	Air Quality	Pros and Cons	References
Wattics	-	wattics	Dublin, Ireland	Win	Hourly	Multi-district and cities	✓				Pros: User-friendly, great for monitoring and analytics with a focus on identifying energy-saving opportunities. Cons: Geared more towards	(wattics)

Lastly, client communication and education are greatly enhanced by the use of white box tools. The detailed outputs from these tools help consultants effectively communicate complex energy concepts to clients. By visualizing energy flows and the impacts of different design choices, these tools bridge the gap between technical energy modelling and client understanding, facilitating the decision-making process.

The time resolution for white box simulation tools in building energy management is crucial for capturing the dynamic interactions of various building components and systems. Typically, these tools offer a range of time resolutions, from annual and monthly down to hourly or sub-hourly intervals. The finer the time resolution, the more detailed the understanding of transient phenomena, such as peak load periods or rapid changes in environmental conditions. For example, an hourly resolution can capture daily cycles of heating and cooling demand, while a sub-hourly resolution might be used to analyze the rapid fluctuations in lighting or HVAC systems due to occupancy changes. Selecting the appropriate time resolution is essential for accurate energy modelling and ensuring optimal energy conservation measures.

As shown in Figure 2.4, each white box simulation tool can be used for a specific time resolution.

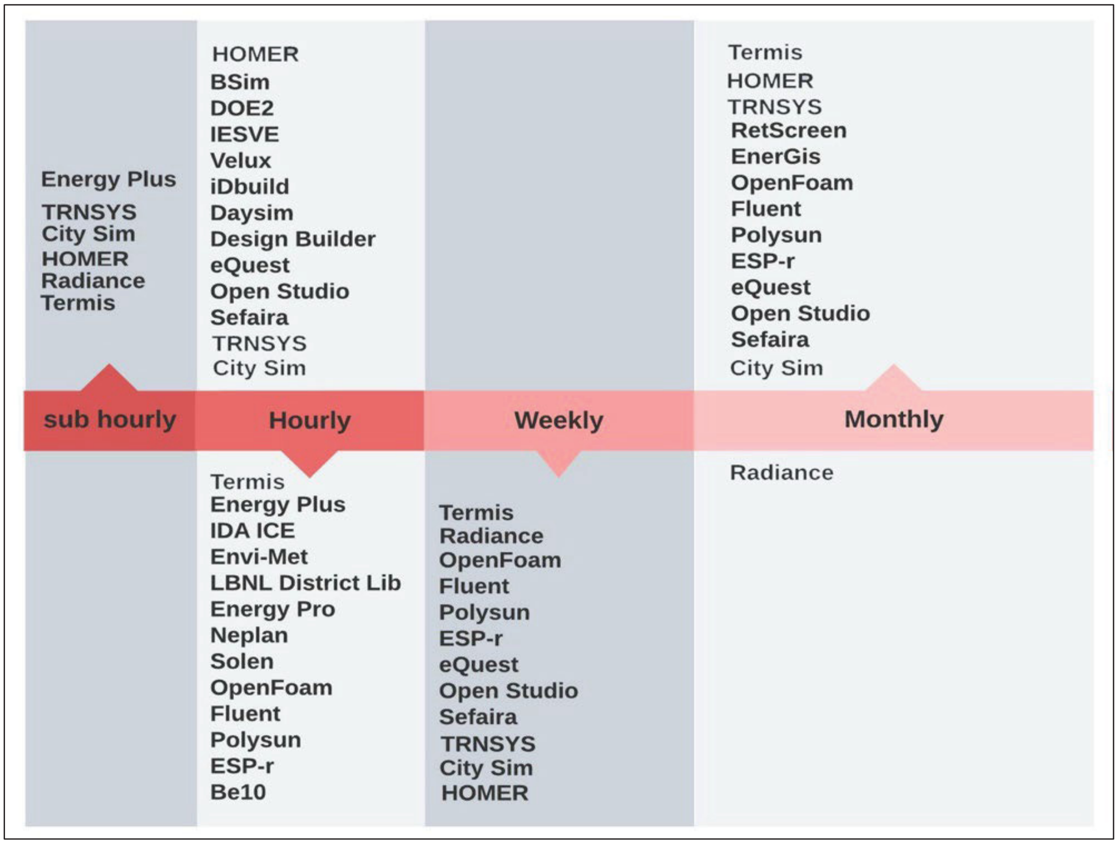

Figure 2.4 The time resolution for white box models]

In the following sections, we describe the most common models and software from Table 1, with the pros and cons.

Among the plethora of white box tools available for building energy management, the selection of “EnergyPlus,” “TRNSYS,” “CitySim,” and “IDA-ICE” is deliberate and strategic. These tools were chosen for their distinct strengths and versatile applications in building energy analysis.

“EnergyPlus” is selected for its exceptional accuracy in simulating building energy performance and its extensive library of building components. It is a robust choice for detailed and precise energy modelling for consultants (Baniyounes et al., 2019). Its broad acceptance in the industry further underscores its relevance. “TRNSYS” is a versatile tool capable of simulating various energy systems and offering customizable component modelling (Hiller et al., 2001). Its adaptability and wide range of applications make it valuable for tackling diverse consulting projects, particularly those involving complex energy systems. “CitySim” is included due to its specialization in urban energy modelling.

The selection of “EnergyPlus,” “TRNSYS,” “CitySim,” and “IDA-ICE” reflects a balanced approach to building energy management consulting, covering a wide range of scenarios and project types. Collectively, these tools offer the precision, versatility, urban focus, and BIM integration required to excel in the field of building energy management.

EnergyPlus

EnergyPlus is an Open-Source (U. d. o. Energy) and comprehensive building simulation software utilized by engineers, architects, and researchers to model energy usage encompassing heating, cooling, ventilation, lighting, plug loads, and water consumption within buildings (Drury B.Crawley, 2001; Sardoueinasab et al., 2018).

EnergyPlus offers integrated solutions, heat balance-based computations, flexible sub-hourly time steps, detailed heat and mass transfer modeling, advanced fenestration and lighting analysis, component-based HVAC modeling, and FMI support (U. d. o. Energy).
EnergyPlus, as illustrated in Figure 2.5, employs a nodal approach that utilizes a one-dimensional conduction transfer function and finite-difference algorithm. Nodal methods are known for their capacity to efficiently address the extensive heat transfer calculations required for building thermal performance, enabling rapid computation (Chen et al., 2022).

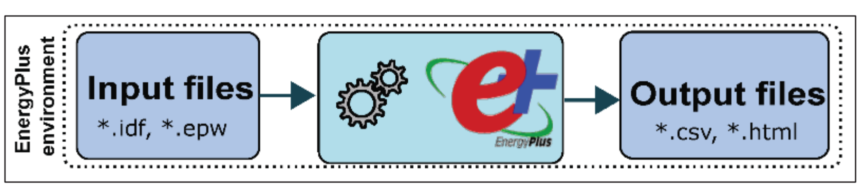

Figure 2.5 General energy plus simulation scheme

EnergyPlus is a versatile simulation tool for analyzing heating, cooling, and ventilation, offering detailed modeling of thermal mass, advanced fenestration, and radiant systems. It supports various HVAC configurations, renewable energy integration, and sustainable strategies like water management and environmental impact assessment. The tool also includes economic analysis and lifecycle cost capabilities, with Modelica and Python integration for custom workflows. Its ability to process diverse and future climate data makes it essential for comprehensive energy performance evaluation (Energyplus).

EnergyPlus is a highly sophisticated white box building energy simulation tool that plays a pivotal technical role in the consulting industry, particularly in building energy management and optimization (Baniyounes et al., 2019). This tool stands out due to its detailed physics-based approach, enabling the precise modelling of building components and systems, including walls, roofs, HVAC systems, and lighting. Consultants rely on "EnergyPlus" for conducting energy audits and predictive analyses of buildings, benefitting from its accuracy in replicating real-world building conditions (Baniyounes et al., 2019). It proves invaluable in retrofitting projects, where understanding the impact of modifications on energy
performance is paramount. Moreover, "EnergyPlus"

TRNSYS

The University of Wisconsin–Madison developed this software package for simulating the behaviour of transient systems using graphical interfaces (A TRaNsient SYstems Simulation Program). TRNSYS can simulate other dynamic systems, including traffic flow and biological processes (Webb et al., 2018). While virtual energy systems are the primary focus of most simulations, TRNSYS is also used for modelling other dynamic systems (TRNSYS).

TRNSYS comprises a kernel engine and a library of 150 modular components. The engine processes input files, performs iterative simulations, and supports regression, matrix inversion, and data interpolation. The component library spans HVAC elements and advanced technologies, allowing customization and scalability (TRNSYS). TRNSYS is particularly effective in consulting contexts, supporting detailed energy audits, retrofitting analysis, HVAC optimization, and compliance verification. Its simulation capabilities, combined with advanced reporting and visualization, enable clear communication of energy performance and facilitate informed decision-making for complex building systems (Tschätsch et al., 2016).

Optimizing ventilation for ambient energy use and heat recovery is key to reducing consumption and maintaining thermal comfort. TRNSYS supports such advanced strategies through its flexible modeling framework, enabling consultants to create tailored simulations that reflect real-world building dynamics. Its versatility makes it a valuable tool for innovative energy management investigations and consulting applications (Hiller et al., 2001). Its module based structure is instrumental for consultants who require precision in simulating and analyzing the energy impacts of potential building modifications and exploring various design scenarios (Korpela et al., 2021; Lipinski et al., 2020; Rashad et al., 2021).

TRNSYS is a foundational tool in building energy consulting, offering a modular and customizable modeling approach. Its flexibility allows consultants to assemble simulations from predefined components, supporting energy audits, retrofitting, and predictive analysis. By accurately modeling building systems, TRNSYS enables precise performance assessments and data-driven decision-making, especially during early design stages (Hiller et al., 2001). TRNSYS's strengths, such as HVAC optimization, compliance verification, and advanced visualization, make it a powerful tool for tailored energy management solutions. Its modular white box framework bridges complex modeling with client-friendly insights, enabling consultants to drive efficient and sustainable building operations (Mueller, 2010).

CitySim

CitySim is a specialized tool for urban energy planning designed to help reduce reliance on non-renewable energy sources and lower greenhouse gas emissions. It provides dynamic, hourly simulations that account for complex interactions like mutual shading and interreflections. By accurately estimating energy consumption and renewable energy potential from individual buildings to entire cityscapes, CitySim supports data-driven decisions for sustainable urban development (Hadziomerovic, 2019).

CitySim is a key tool in urban energy consulting designed for complex, multi-building environments. It enables precise energy analysis, supports audits and retrofit simulations, and facilitates data-driven recommendations. By accounting for climate and environmental factors, CitySim helps assess building performance within urban contexts. Its scalability, technical customization, and compliance support make it adaptable to diverse project needs. With strong visualization and reporting features, CitySim bridges technical insights and client understanding, solidifying its role as an advanced white box solution in energy consulting (Perez et al., 2011).

CitySim is a dynamic urban energy simulation tool that estimates energy consumption and renewable potential at building and district scales. It performs hourly simulations that account for seasonal demand variations, mutual shading, HVAC performance, and occupant behavior. Ideal for energy audits and retrofit planning, CitySim integrates local climate data to support regulatory compliance. Its adaptability, scalability, and advanced visualization capabilities make it a key asset for sustainable urban energy management (Perez et al., 2011).

IDA ICE is a versatile building-level simulation tool utilizing the neutral model format (NMF) for equation-based simulations. It boasts an intuitive graphical user interface, the capability to import industry foundation class (IFC) models, and the ease of extending functionality by creating new components (Kräuchi et al., 2014). While district-level modeling has been limited, recent advancements, particularly in low-temperature bidirectional district heating, are expanding its scope through models that include heat supply systems, distribution pipes, pumps, and borehole heat storage (Allegrini et al., 2015). IDA ICE is a cornerstone tool in energy consulting, known for its high-precision dynamic simulations at the building level. It enables detailed modeling of systems and components, offering hourly outputs that capture energy consumption fluctuations. The tool supports comprehensive evaluations of energy-saving measures and climate impact assessments and offers parametric analysis for project-specific customization, making it indispensable for building energy optimization simulations for specific projects, optimizing building energy performance. IDA ICE ensures compliance with energy standards and enhances client communication through advanced visualization and reporting features. IDA ICE is a technically advanced white box building energy tool that equips consultants with the precision and versatility needed for effective consulting practices in building energy management (Martin, 2017).

With features for both steady-state and dynamic simulations, IDA-ICE facilitates the exploration of energy-saving measures through retrofitting and renovation projects (Mohammadusman Doddamani, 2023; Moser et al., 2019).

Figure 2.6 demonstrates how white box models can derive hourly cooling and heating load patterns from input files, including domestic hot water profiles, domestic electricity profiles, building stock archetypes, and U-values. These hourly profiles can then serve as inputs for black box models or assist in determining the optimal balance between supply and demand.

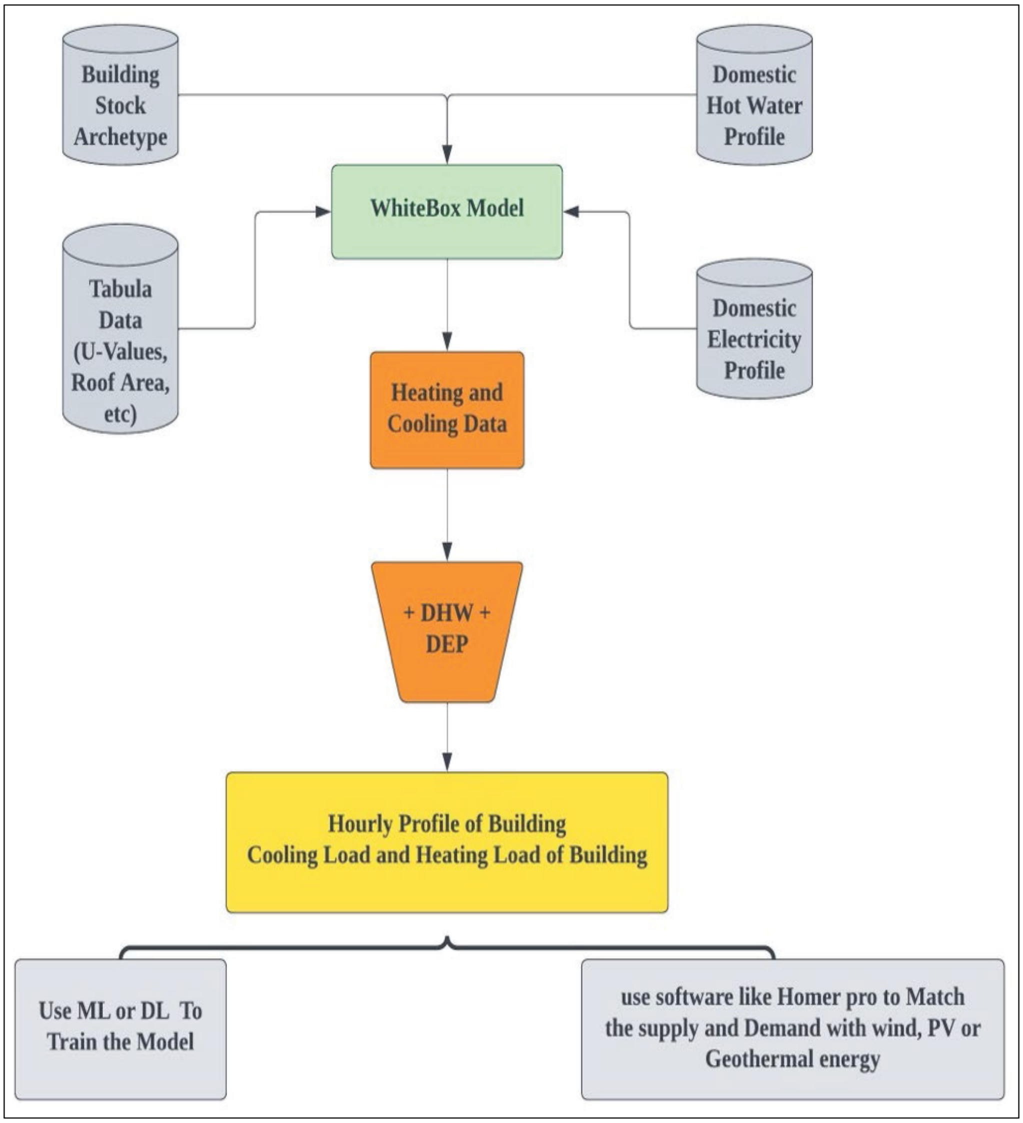

Figure 2.6 Different outputs using white box (like IDA-ICE or TRNSYS) and black box models (ML or DL)

The white box models can extract energy, thermal, daylight, and air quality from these four outputs, as shown in Figure 2.7.

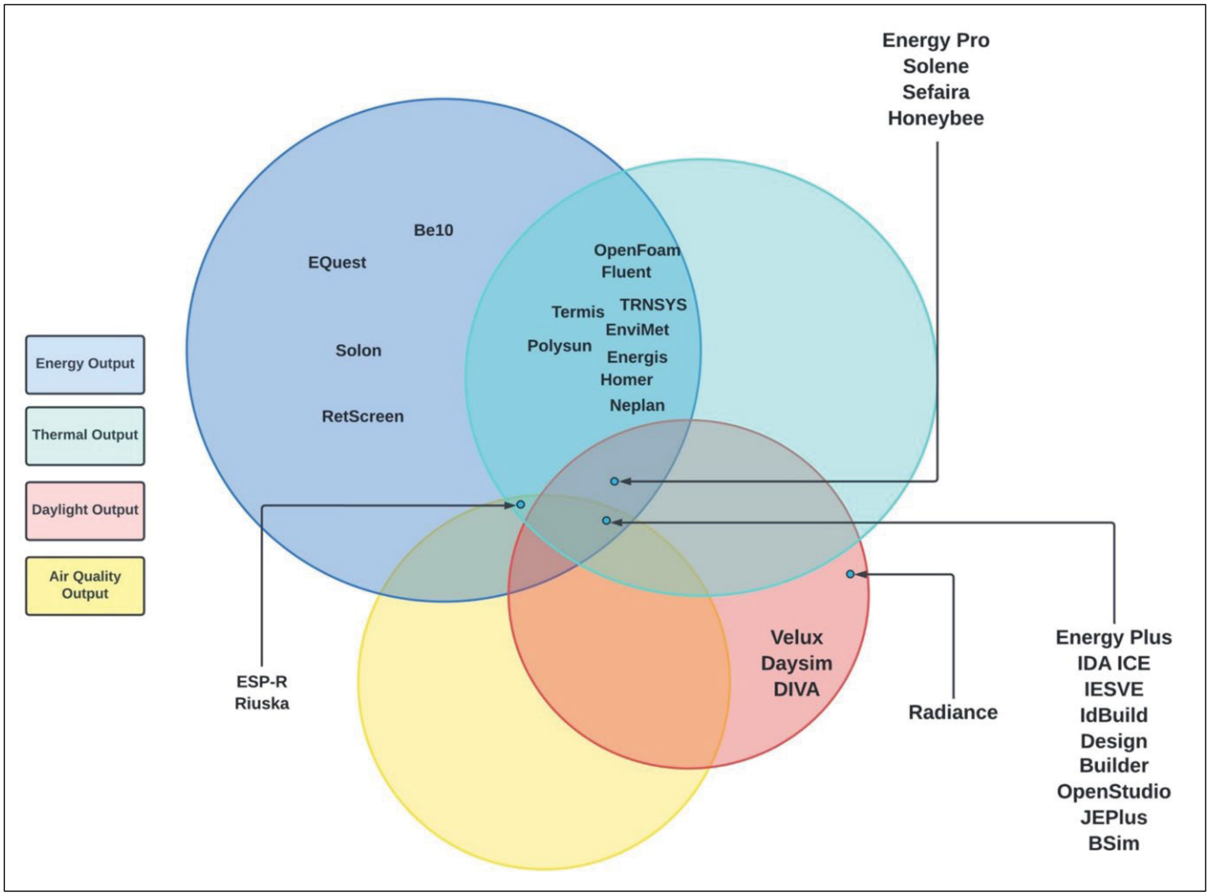

Figure 2.7 Outputs for white box energy simulation tool

2.4.1.2 Black Box Models

Data-driven modelling (DDM) uses external data to define configurator model components and inject them into the simulations. In different application domains, data-driven models are becoming increasingly popular thanks to progress in computational intelligence and machine learning techniques. The input/output data from real-world systems are used to develop data-driven models rather than analytical or numerical models. Modelling based on data-driven inputs and outputs is described in control and systems engineering by collecting inputs and outputs, choosing a model category, estimating model parameters, and confirming the accuracy of the estimated model.

From simple linear regression (Ciulla & D'Amico, 2019) to more elaborate deep learning methods (Mocanu et al., 2016), energy calculations based on data-driven methods can be performed at various levels of intricacy. The literature (Sun et al., 2020; Wei et al., 2018) contains reviews of other methods for data-driven predictions.

Data-driven techniques can be grouped based on their statistical models (such as Support Vector Machine models and Artificial Neural Network models), the type of data they use (empirical or pre-simulated), and the variables they predict. Furthermore, these methods can be categorized according to their specific applications, including design, peak load estimation, fault diagnosis, and system tuning.

Data-driven models eliminate the need for building thermal balance equations, reducing or eliminating the requirement for detailed physical building information (Chen et al., 2022). Through mathematical techniques, data-driven models uncover the hidden connections between output variables, such as building energy consumption, and input variables, like weather, building details, occupant behaviour, and equipment schedules. These methods readily apply to buildings that lack comprehensive physical parameters, such as those in the construction phase.

Based on Figure 2.8, black box models start with loading building and meteorological datasets. Public (Open Dataset by Ministry of Housing, Communities & Local Government; US Government Open Data Platform) or private (KTH University Stockholm) datasets can be used to access building datasets. On the other hand, government meteorological platforms (Historical Climate Data by Government of Canada) provide access to meteorological datasets. Pre-processing the data is the next step. There is a difficulty with different algorithms, as they make assumptions about your data and may require additional transformations to be applied. However, it is also possible for algorithms to deliver better results without preprocessing when all the rules have been followed and the data have been prepared.

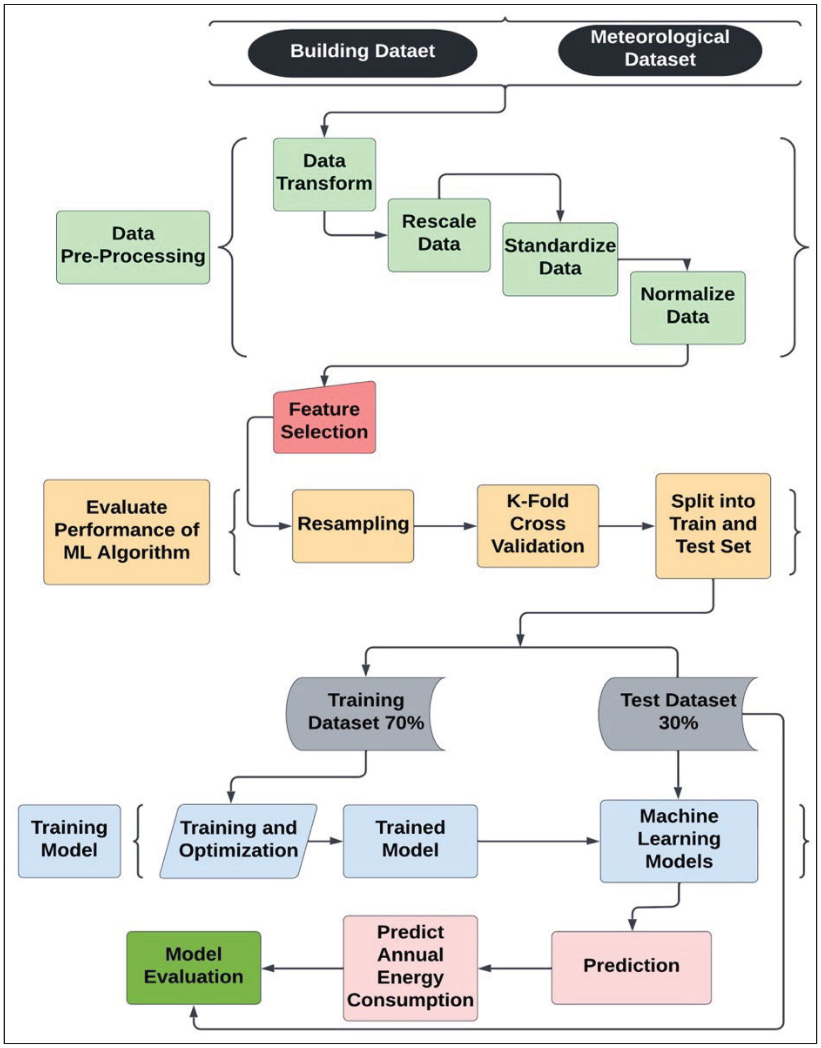

Figure 2.8 The procedure of black box models

The data should be viewed in many ways, and then a handful of algorithms should be applied to each view. The steps for Pre-processing Data include Transforming Data, Rescaling Data, Standardizing Data, and Normalizing Data. The Data features significantly influence a machine learning model's performance (Brownlee, 2016). Features that are unrelated or only partially relevant can have a detrimental effect on a model's performance. The feature selection process involves selecting those features in the data most likely to contribute to a prediction or output. Many models, especially linear algorithms like linear regression and logistic regression, can become less accurate when irrelevant features are present in the data (Brownlee, 2016).

Evaluating an algorithm's performance on unseen data is crucial. The most effective method is to make predictions on new data with known outcomes, followed by statistical resampling methods for accurate performance estimation. Machine learning algorithms are assessed using different training and testing datasets, where an algorithm is trained on one part of the dataset, predictions are made on the other, and these predictions are compared against expected outcomes. The black box model is valued for its simplicity and reliance on actual performance data, suitable for benchmarking across multiple structures. However, it needs more detailed insights into energy inefficiencies and conservation measures, and its accuracy depends on the quality and relevance of historical data. Significant deviations in the data, like major building retrofits, can reduce its predictive power. While useful for Monitoring and Verification (M&V) and forecasting, the black box model does not provide the in-depth analysis of white or grey-box models but complements them in building energy modelling.

The current review summarizes black box models in Table 2.2 below, showing detailed information on the most well-known. All black box models, based on machine learning or programming codes, can be used on Windows, Mac, or Linux because these simulation tools do not depend on the operating system.

Black box tools in building energy management have emerged as a cornerstone in consulting, offering empirical, data-driven solutions pivotal for quick decision-making and effective energy management strategies. These tools, grounded in statistical and machine learning models, are employed across various consulting domains, including operational optimization and predictive maintenance (Hopfe et al., 2017). Operational optimization and energy profiling heavily rely on black box tools within the consulting realm (Hopfe et al., 2017). By analyzing historical energy usage data, these tools discern patterns and anomalies, aiding consultants in offering recommendations to optimize energy consumption and minimize costs (Deb &

Schlueter, 2021). The energy profiling process, which is about understanding a building's energy consumption patterns, is significantly enhanced by black box models, thanks to their proficiency in processing extensive datasets and pinpointing trends.

In predictive maintenance and fault detection, black box tools show their strengths in building management systems. They utilize historical data to anticipate maintenance needs, averting equipment failures and downtime. These tools are also pivotal in detecting faults in building systems early on before they escalate into serious issues. For retrofit analysis and scenario simulation in retrofit projects, black box tools simulate various scenarios to shed light on potential energy savings and return on investment. Their rapid processing of different scenarios assists consultants in suggesting the most cost-effective and energy-efficient retrofit measures.

Real-time energy management and demand response strategies extensively leverage black box tools. These tools enable consultants to offer instant recommendations for energy optimization and are integral in demand response strategies, where buildings alter their energy usage in response to external cues like peak demand periods or fluctuations in energy prices (Deb & Schlueter, 2021).

The adaptability of black box tools allows consultants to provide tailored solutions for various building types and sizes, which is especially beneficial in buildings with unpredictable energy usage patterns or where in-depth physical modelling is impractical. Client reporting and visualization are other areas where black box tools excel. Equipped with advanced reporting and visualization features, these tools help consultants communicate complex energy data to clients in an understandable manner, converting vast data into comprehensible reports and graphs. Moreover, consultants employ black box tools for benchmarking the energy performance of buildings against counterparts or industry standards. This benchmarking is crucial for pinpointing areas for energy performance enhancement and guiding strategic planning for energy efficiency improvements.

Table 2.2 Black_Box Models

Software/Code	Version	Developer	City, Country	Language	System Boundary	Available Outputs				Pros and Cons	References
						Energy	Thermal	Electrical	Lighting		
Open IDEAS	-		Paris, France	Modelica, Motoko, Python	-		✔	✔		Pros: Integrates with the Modelica language, allowing for flexible, physics-based modelling of building energy systems. Cons: Requires knowledge of the Modelica language.	(Baetens et al., 2015; Leuven, 2015)
TEASER	0.7.7	RWTH Aachen University	-	Python, Modelica	Single-Family and Multi-Family House	✔				Pros: Quick setup of building energy models for urban-scale simulations. Cons: Does not have the depth and detail needed for fine-tuned building-specific energy analysis.	*(RWTH Aachen University,TEASER; TEASER*
CityLearn	2.1.0	Intelligent Environments Laboratory	Berkeley, CA, USA	Python	City	✔				Pros: Designed to facilitate multi-agent reinforcement learning. Cons: Requires knowledge of reinforcement learning techniques.	*(Intelligent Environments Laboratory City*
PyCity	0.3.3	RWTH Aachen University	Aachen, Germany	Python	Neighborhood, Districts	✔	✔			Pros: Python-based tool that offers flexibility and integration with other Python libraries and tools. Cons: Python proficiency is needed.	(Institute for Energy efficient Buildings and
RC Building Simulator	-	Prageeth Jayathissa, et al.	-	Python	Single-Family and Multi-Family House	✔			✔	Pros: Simplifies the process of building thermal modelling using RC models. Cons: Oversimplification may miss out on more complex interactions.	(pjayathissa)
Open energy modelling framework	1.0	Oemof developer team	Open source	Python	-	✔				Pros: An open-source framework that can be tailored to various energy system modelling needs. Cons: Might require more effort to set up and customize compared to out-of-the-box solutions.	*(Open Energy Modelling Framework*
OCHRE	0.8.4	NREL	Chicago, IL, USA	Python	-	✔				Pros: Targets optimal control and hardware-in-the-loop simulation. Cons: It may not be as widely applicable or supported as more established tools.	(Blonsky et al., 2021; NREL)
Building Energy Platform	-		-	Python, Java	Multi Districts, city, neighborhood, Districts, single-family house, multi-family house	✔				Pros: Potentially integrating various data sources. Cons: The platform may depend on the availability and quality of data inputs for effective energy management.	*(Building Energy Platform,marcelo bastos, 2014)*
Building automation energy data analytics (BAEDA)	-	Team from Polytechnic of Turin University	-	Python	Single-Family House, Multi-Family House	✔		✔		Pros: Designed to analyze data from building automation systems to improve energy efficiency. Cons: May require complex integration with existing building automation systems and substantial data processing capabilities.	(Pinto)

These four tools, Linear Regression, Support Vector Machines (SVM), Random Forest, and Deep Neural Networks (DNN), have been selected for detailed explanation among various other options in the domain of black box models for BEM. This choice has been made due to their wide adoption and effectiveness in handling diverse building energy optimization tasks. Linear Regression provides a fundamental understanding of relationships between variables and serves as a baseline model. SVM, known for its robustness in handling complex data, offers excellent classification capabilities. Random Forest is selected for its ensemble learning approach, which enhances predictive accuracy. DNN, as a deep learning model, has shown remarkable success in capturing intricate patterns and nonlinear relationships in building energy data. These four tools collectively represent a comprehensive spectrum of modelling approaches, making them crucial for consultants and professionals engaged in building energy management, thus warranting in-depth exploration.

Linear Regression (LR)

LR has been widely used in many fields since it has good predictive performance and is simple. Linear and nonlinear regression methods are both regression methods (Géron, 2022). Linear regression, when applied as a black box model in building energy management and the broader energy industry, presents both advantages and challenges. At its core, linear regression seeks to model the relationship between one or more independent variables (predictors) and a dependent variable (often the energy consumption in this context). As a black box approach, it emphasizes prediction accuracy over the interpretability of the underlying relationships. This predictive nature allows stakeholders in the energy industry to make quick, data-informed decisions about energy use and demand forecasting without necessarily understanding the intricate physical processes behind it.

Linear regression serves as a foundational analytical method within the energy management discipline, offering a straightforward and computationally efficient approach to model and forecast energy consumption. Its utility in the energy sector is underscored by its ability to inform resource distribution, enhance operational efficiency, and discern consumption trends. Energy managers frequently deploy linear regression for load forecasting and demand-side management, as well as for crafting predictive models of energy usage that facilitate swift, data-driven decision-making. While linear regression is esteemed for its simplicity and ease of use, which are advantageous for prompt analyses, it is also important to recognize its limitations in encapsulating the complex, nonlinear interdependencies typical of advanced energy systems. This recognition underscores the necessity for an eclectic array of analytical tools to comprehensively address the multifaceted nature of energy system modelling and management (Harish & Kumar, 2016). Linear regression is well-suited for applications involving linear or nearly linear relationships between variables, such as baseline energy consumption modelling, essential forecasting, and analyzing the influence of factors like outdoor temperature on energy usage. However, it may not effectively capture complex nonlinear relationships found in intricate building systems (Ciulla & D'Amico, 2019).

Technically, this tool models relationships between independent variables (predictors) and a dependent variable, often representing energy consumption. Consulting is frequently employed to create baseline energy consumption models, enabling consultants to establish reference points for energy management. Additionally, Linear Regression supports straightforward energy consumption forecasting, aiding consultants in short-term predictions and energy-efficient planning. Moreover, it helps identify critical factors affecting energy usage, such as temperature or occupancy, providing valuable technical insights for consultants. Its simplicity and computational efficiency make it a good choice for rapid data-driven decision-making. However, it is essential to acknowledge that Linear Regression may have limitations in capturing complex, nonlinear relationships, which are prevalent in intricate building systems. Nonetheless, its technical advantages position it as a valuable tool for consultants seeking quick and practical insights into building energy management within the consulting industry.

Support Vector Machine

Support vector machine (SVM) comprises a range of supervised learning techniques employed for tasks like classification, regression, and identifying outliers (Library). Support vector machines have several advantages, such as effectively handling high-dimensional spaces. Furthermore, this approach remains efficient even when dealing with datasets with greater dimensions than the number of samples. Additionally, the decision function relies on a subset of training points known as support vectors, ensuring memory efficiency. Finally, SVM's ability to solve nonlinear problems is one of its most essential capabilities (Chen et al., 2022; Library).

Support Vector Machines (SVMs) stand out in the energy management sector for their robust predictive capabilities, particularly within building energy optimization. As a class of powerful black box models derived from machine learning, SVMs are adept at classification and regression tasks, which are essential for analyzing and interpreting complex energy datasets. Their application in energy management is particularly valuable for forecasting consumption patterns, identifying irregularities in energy usage, and categorizing different operational states of energy systems. SVMs are prized for their proficiency in handling multi-dimensional datasets and their capacity to elucidate intricate nonlinear relationships frequently occurring in building energy systems. Utilizing the kernel trick, SVMs can transform nonlinear data distributions into a format amenable to analysis, which is particularly beneficial for modelling the dynamic interactions within building energy systems. By providing accurate and refined models of energy behaviours, SVMs contribute significantly to enhancing energy efficiency measures and optimizing operational processes, solidifying their role as indispensable assets in the strategic toolkit of energy management professionals (Brownlee, 2016). Derived from machine learning, SVMs excel in classification and regression tasks by identifying optimal hyperplanes in high-dimensional spaces to classify or predict data points. In consulting, these tools offer several technical advantages. They are instrumental in predicting energy consumption patterns, enabling consultants to forecast and understand energy usage, a fundamental requirement in consulting.

Additionally, SVMs excel in anomaly detection, helping identify irregularities or inefficiencies in energy use and contributing to efficient energy management.

Moreover, they can classify building operational states, offering insights into building performance and facilitating data-driven decision-making. Their proficiency in handling high-dimensional data and capturing complex nonlinear relationships between variables is particularly relevant in building energy dynamics' intricate and multifaceted domain. Overall, SVMs are a technically powerful asset for consultants in building energy management, aligning with the industry's complex and dynamic nature (Magoulès & Zhao, 2016). SVMs, especially with nonlinear kernels, can capture intricate relationships in data. They are suitable for predicting energy consumption patterns, detecting anomalies in energy usage, and classifying the operational states of building systems. Their ability to handle high-dimensional data can be beneficial when integrating multiple sensors and systems (G. Zhang et al., 2022).

Random Forest

Random forest algorithms combine bagging and feature randomness to create uncorrelated forests of decision trees, which is an extension of the bagging method. Three primary hyperparameters need to be set before training a random forest algorithm. There are three main factors: node size, trees, and feature samples.

The Random Forest algorithm offers advantages and challenges for classification and regression tasks. It notably reduces overfitting, a common issue in decision trees, by averaging uncorrelated trees, making it popular for accurate regression and classification tasks among data scientists (Zekić-Sušac et al., 2021). The Random Forest method is crucial in the energy industry because it provides accurate predictions while mitigating overfitting, handling missing data, and determining feature importance. Its flexibility in addressing regression and classification tasks, scalability for large datasets, and capacity to capture nonlinear relationships make it a versatile tool.

It plays a vital role in energy diagnostics, enabling energy consumption predictions, fault detection, and identification of key drivers for energy use. Random Forest is invaluable for optimizing energy performance and improving operational efficiency in the energy sector. A critical challenge of Random Forest is its time-consuming nature, as it processes data slowly when computing each decision tree. It also consumes more resources because it handles larger datasets, potentially limiting data storage resources. Moreover, it is more complex than a single decision tree, making predictions less straightforward to interpret (Nazarenko et al., 2019).

Leveraging ensemble learning, Random Forest combines multiple decision trees to provide robust predictions and insights. Its key advantages include the reduction of overfitting, adaptability for various tasks (classification and regression), the ability to handle missing data, straightforward determination of feature importance, and the capture of complex relationships between variables. These qualities make it invaluable in consulting, which finds applications in predicting energy consumption, detecting system faults, and assessing feature importance. Its versatility, high predictive accuracy, and capability to handle diverse tasks make it an indispensable tool for consultants engaged in building energy management and optimization (Magoulès & Zhao, 2016).

The Random Forest method is an ensemble learning technique widely recognized for its effectiveness in energy management. By leveraging multiple decision trees to generate predictions and classify data, it excels at handling the multifaceted nature of energy systems, from predicting energy demand and consumption to optimizing distribution and detecting inefficiencies. Its ability to manage large datasets with numerous variables captures the complex interplay between usage patterns and environmental factors while also identifying key predictors of energy performance for targeted energy-saving strategies. Furthermore, its strong predictive accuracy and resistance to overfitting are invaluable for ensuring reliable decision making in energy efficiency initiatives, renewable integration, and load forecasting.

This method's comprehensive analytical strengths make it an essential component of data-driven decision-making in energy management, providing actionable insights for enhancing system reliability and sustainability (Wang et al., 2018).

Deep Neural Networks

Deep Neural Networks (DNNs- Figure 2.9), applied as black box models in building energy management, excel at learning from extensive data, making them invaluable in this field. Their architecture comprises multiple interconnected layers capable of automatically extracting complex data patterns. This inherent ability allows DNNs to model building energy systems' intricate and nonlinear dynamics effectively. Whether the task involves energy consumption prediction, HVAC optimization, or anomaly detection, DNNs consistently outperform traditional models, adapting well to diverse building types, systems, and external factors like weather conditions (Zekić-Sušac et al., 2021).

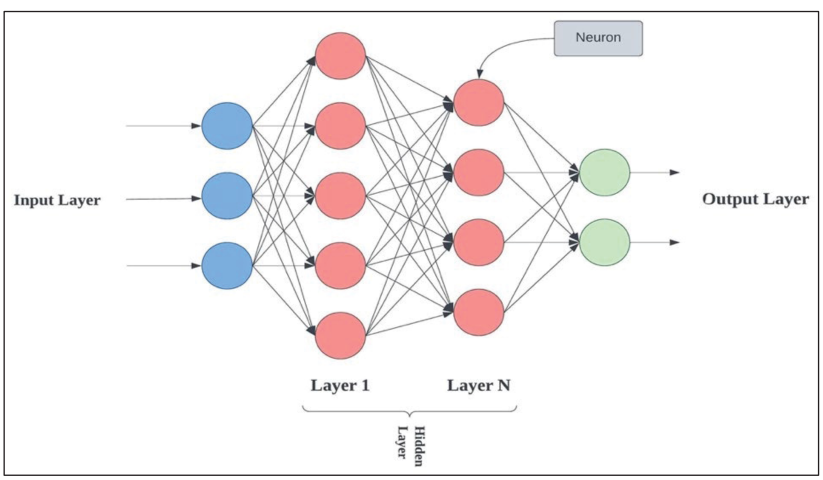

Figure 2.9 A Deep Neural Network with N Hidden Layers

Deep Neural Networks (DNNs) have become indispensable in the energy industry due to their capacity to handle large and complex datasets. In building energy management, DNNs are crucial in predicting energy consumption, optimizing HVAC systems, and detecting anomalies, often outperforming traditional models.

Their significance lies in their ability to capture intricate patterns and relationships within energy systems, making them valuable tools for improving energy efficiency. However, their black box nature can be a challenge, making it challenging to interpret how they arrive at their predictions. This opacity may pose issues for stakeholders seeking to uncover energy inefficiencies or meet regulatory transparency requirements. DNNs also require substantial amounts of labelled data, which can be limited in building energy contexts. Additionally, setting up and fine-tuning DNNs can be a complex task. Despite these hurdles, the substantial predictive power of DNNs and ongoing research in model interpretability ensure their continued prominence in the future of the energy industry, where data-driven decision-making and sustainability are paramount (Magoulès & Zhao, 2016; Sze et al., 2017).

Deep Neural Networks (DNNs) have emerged as a powerful tool in building energy management consulting practices. DNNs, consisting of multiple interconnected layers of nodes, excel at automatically extracting intricate features and patterns from data. This inherent capability makes them well-suited for modelling the complex and often nonlinear dynamics of building energy systems. Consultants leverage DNNs for various applications, including energy consumption forecasting, optimizing HVAC operations, and detecting anomalous energy usage patterns. DNNs have demonstrated superior performance compared to traditional machine learning models, adapting to diverse building typologies, systems, and external factors such as weather patterns (Marino et al., 2016). However, DNNs also need help in consulting. Their black box nature, while providing accurate predictions, lacks interpretability, making it challenging to uncover the underlying physical relationships behind energy behaviours. This opacity can be problematic when stakeholders must understand the root causes of energy inefficiencies or meet regulatory transparency requirements. Training DNNs requires substantial amounts of labelled data, which may only sometimes be readily available in building energy contexts. The selection of the right architecture, hyperparameter tuning, and avoiding overfitting demand careful consideration, making the application of DNNs non-trivial (Chen et al., 2017).

Despite these challenges, DNNs' exceptional predictive capabilities and ongoing research in model interpretability ensure their continued prominence in the consulting landscape of building energy management. They provide consultants with a powerful tool to analyze and optimize energy systems, ultimately leading to more efficient and sustainable building practices.

2.5 Web Based Models

Web simulation tools can be categorized into white box and black box, similar to standalone simulation tools. Web-based interfaces for building energy systems have gained popularity due to their cost-effectiveness, compatibility across platforms, and ease of maintenance. These interfaces offer advantages such as cloud hosting and computing, which prevent data loss and facilitate data exchange.

In recent years, many tools have adopted web interfaces to visualize energy data, enabling benchmarking and detailed analysis of simulation results for urban buildings. These web-based tools can create 3D energy models of urban buildings and overlay colour codes to represent energy performance levels, allowing filtering by size, type, location, and building system for in-depth analysis (Sola et al., 2018).

These applications' simple interface and quick outputs dramatically lighten the load for users. Furthermore, it has several significant benefits for users. These tools are automatically updated by updating the server's source code, so users do not need to update them themselves. In addition, they are more suitable for teams since they make it easier for team members to share the project information and have little risk of losing project data. Finally, they may use cloud computation more effectively. Cloud computing technology can reduce calculation time and system processing power, which is critical to building performance simulations (Grove, 2009; Sola et al., 2018). Various time resolutions, from one hour to a year, are available through the web tools, as shown in Figure 2.10.

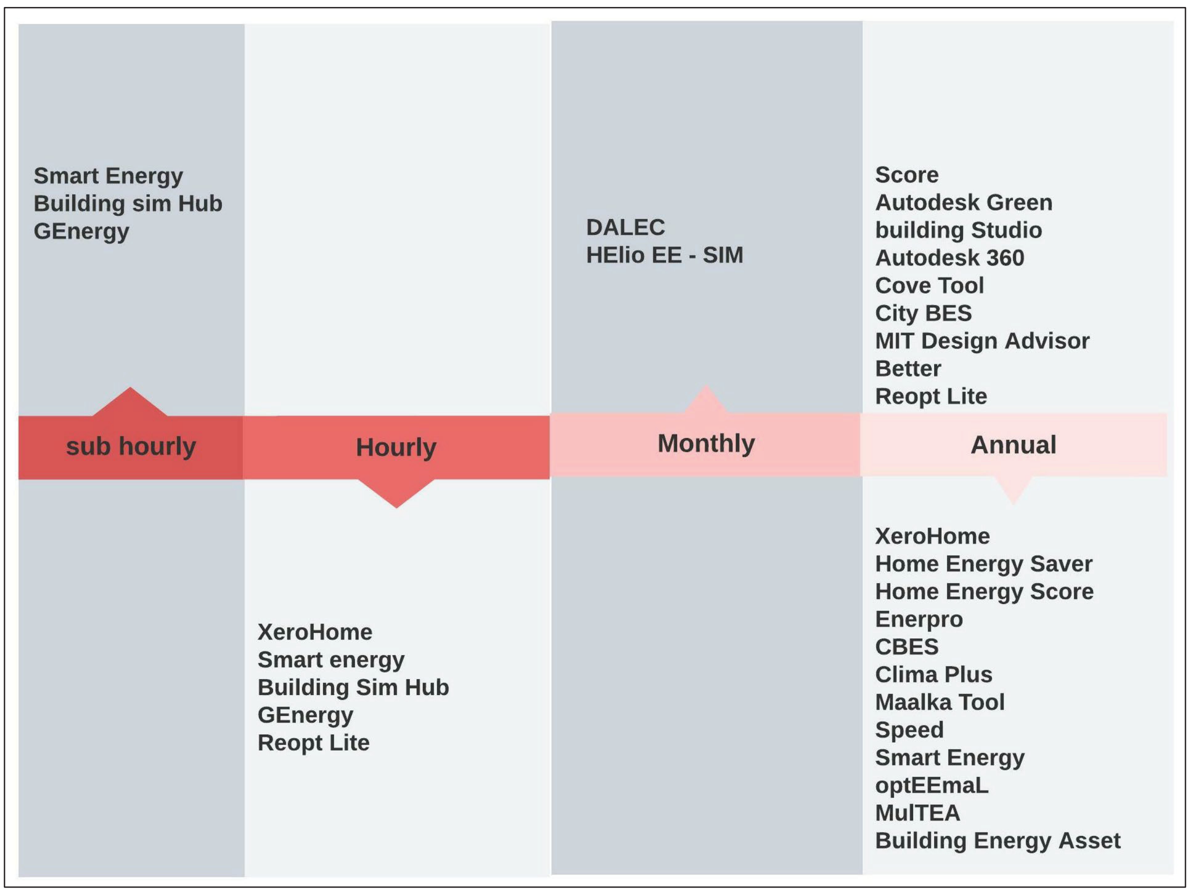

Figure 2.10 Time resolution for Webtools

Compared with standalone simulation tools, web-based simulation tools also have several disadvantages. The GUI of browser-based applications (such as HTML) is limited compared to desktop applications, meaning that a single function might be repeated across multiple pages

(Prokhorenko et al., 2016). Users must navigate several pages and follow a sequence pattern to perform a simple calculation or analysis (Claudio, 2020; Dissanayake & Dias, 2017; Prokhorenko et al., 2016). These applications provide slower response times because of work wait patterns and standby modes that limit their interaction on page refreshing (Claudio, 2020; Dissanayake & Dias, 2017; Gupta & Gupta, 2017), and the whole process is slowed down by longer response times.

A fast and stable internet connection is required for these applications to function correctly and ensure fast data transmission (Rak et al., 2019; Rusmardiana et al., 2018; Vlčková et al., 2017). Security issues, such as the risk of reverse engineering and hijacking of source code (Ismanto & Salman, 2017), lack of management, maintenance, and modification, network connectivity limitations, limited access to local resources, and difficulties in debugging (Hokamp, 2018) are the most reported disadvantages.

Web-based tools in building energy management have become integral to consulting practices, offering a range of technical capabilities that support consultants in their energy analysis and decision-making processes. These tools are accessible via web interfaces, making them convenient and cost-effective for consultants. They have a significant impact on consulting practices by facilitating data visualization and benchmarking, offering simulation and modelling capabilities, utilizing cloud hosting and data exchange for enhanced data security, supporting collaboration and client communication, providing comprehensive energy performance analysis, and ensuring accessibility and flexibility. In summary, web-based tools enhance consulting practices by empowering consultants to provide data-driven insights, optimize building energy systems, and make informed decisions for energy efficiency improvements while effectively engaging clients in the process (Ang et al., 2022).

Table 2.3 summarizes the software’s features, such as developer, zone analysis level, building type, available outputs, and modelling approach for web tools.

Table 2.3 Web Tool Simulation Tools

Software/ Code	**Version**	**Developer**	**City, Country**	**Black-Box/White-Box**	**System Boundary**	**Available Outputs**				**Timeframe**	**Pros and Cons**	**Reference**
						Energy	**Thermal**	**Electrical**	**Daylight**			
Xerohome	23503	Mudit Saxena, Peter Mayostendorp, Inderdeep Dhir	CA, USA	White-Box Model	Single-Family House and Multi-Family House	✔				Dynamic	Pros: Offers detailed modelling of home energy efficiency. Cons: Only for US.	(MUDIT SAXENA)
Home Energy saver	-	Berkely Lab	-	Black-Box Model	Single-Family House and Multi-Family House	✔				Annual	Pros: Provides homeowners with personalized energy use assessments and improvement recommendations. Cons: Only for US.	(*Home Energy Saver*)
Home Energy score	-	U.S. Department of Energy	USA	Black-Box Model	Single-Family House and Multi-Family House	✔				Annual	Pros: Gives a quick and straightforward assessment of a home's energy efficiency and potential improvements. Cons: Simplified scoring may not reflect the complexities of individual homes' energy dynamics.	(*Home Energy Score*)
Enerpro (The Energy Profile Tool)	9.2.1	EnerSys Analytics Inc. and XModus Software Inc.	Vancouver, BC, Canada	White-Box Model	Single-Family House and Multi-Family House	✔				Annual	Pros: Allows for quick benchmarking of a building's energy performance against similar structures. Cons: May not provide detailed suggestions for energy improvements.	(*EnerPro*)
Senapt		Senapt Team	UK	-	Single-Family House and Multi-Family House	✔					Pros: Can assist in monitoring and managing energy consumption. Cons: May require technical expertise.	(*Senapt*, 2020)
CBES	2.0	CIPSEA		Black-Box Model	Multi-Districts and City Scale	✔				Annual	Pros: Provides quick energy efficiency assessments. Cons: The tool's recommendations may be less specific than those obtained from a detailed analysis.	(Lee et al., 2015)
ClimaPlus	Climasplus 2020		MA, USA	White-Box Model	Single-Family House and Multi-Family House					Annual	Pros: Focuses on climate data analysis to inform building design and retrofit strategies for energy efficiency improvements. Cons: Its use may be limited if climate data integration is not a central component of the energy management strategy.	(*Clima Plus*)
Maalka Tools		Maalka Inc., NY	NY, USA	White-Box Model	Single-Family House and Multi-Family House	✔	✔			Annual	Pros: Offers a platform for managing sustainability metrics and energy performance data. Cons: May require significant data input.	(*Maalka Tool*)
Speed	2021		Wasliiiigton, DC, USA	White-Box Model	Single-Family House and Multi-Family House				✔	Annual	Pros: Designed for rapid energy modelling. Cons: The speed of analysis might come at the expense of model depth and accuracy compared to more detailed simulation tools.	(*Speed simulation platform*)
Smart energy	3.0.1		USA	White-Box Model	Single-Family House and Multi-Family House	✔				Hourly, Sub Hourly	Pros: Enables detailed analysis and optimization of energy consumption, aiming to improve overall building energy efficiency. Cons: Its effectiveness greatly depends on the availability and granularity of energy consumption data fed into the system.	(*Smart Energy*)
OptEEmAL	2019	European Union	Spain, Germany	White-Box Model	Single-Family House and Multi-Family House	✔				Annual	Pros: Offers a platform for optimizing energy-efficient building retrofit plans using integrated project delivery methods, which can enhance collaboration and efficiency. Cons: May require complex data and modelling inputs.	(García-Fuentes et al., 2019; Hernández et al., 2017)

Software	Version	Developer	City, Country	Black Box/White Box	System Boundary	Energy	Thermal	Daylight	Air Quality	Time Frame	Pros and Cons	References
Building Energy asset score	2014	Us department of Energy	USA	Black-Box Model	Single-Family House and Multi-Family House	✔			✔	Annual	Pros: Developed by the U.S. Department of Energy, it assesses the energy efficiency of building assets and provides a score, making it useful for benchmarking and understanding potential improvements, open source. Cons: Primarily focused on the inherent energy performance of the physical building assets, which may not account for operational variables.	(Lee et al., 2015)
CityBES	-	Lawrence Berkeley National Lab, under the Laboratory Directed Research and Development	Berkeley, MA, USA	White-Box Model	Multi-Districts and City Scale	✔		✔		Annual	Pros: A tool designed for urban-scale analysis, it helps in evaluating energy savings and carbon reduction strategies for city-wide building stocks. Cons: Its urban focus might make it less applicable for individual building projects or more detailed energy system design.	(Hong et al., 2016)
Autodesk Green Building Studio	2023	Autodesk	San Rafael, CA, USA	Black-Box Model	Multi-Districts and City Scale	✔	✔		✔	Annual	Pros: Integrates with other Autodesk design software, enabling seamless energy analysis within the design process, useful for architects and designers. Cons: As part of a suite of design tools, it may not have the depth of standalone energy simulation software.	(AutoDesk, 2015)
Autodesk insight 360	2023	Autodesk	San Rafael, CA, USA	Black-Box Model	Multi-Districts and City Scale				✔	Annual	Pros: Provides cloud-based energy modelling that is integrated with BIM (Building Information Modelling), offering user-friendly insights into the energy and environmental design of buildings. Cons: Might require a subscription to the Autodesk suite, and its simplified interface may not offer the granularity needed for complex engineering analyses.	(AutoDesk)
BuildingSimHub	2017	Us department of Energy	France	White-Box Model	Multi-Districts and City Scale	✔				Hourly, Sub Hourly	Pros: Offers a cloud-based simulation platform that streamlines the building energy modelling process, making it accessible for collaboration across different stakeholders. Cons: Being cloud-based, it may face limitations with data security concerns or require a stable internet connection for optimal use.	(Buildsim)
Rescheck web	-	Us department of Energy	USA	White-Box Model	Multi-Districts and City Scale	✔				-	Pros: Provides a straightforward method for demonstrating building energy code compliance, with a focus on residential buildings, and is a free web-based tool offered by the U.S. Department of Energy. Cons: While useful for code compliance, it may not offer the detailed analysis required for optimizing energy consumption beyond the minimum code requirements.	(*ResCheck Web*)
Cove Tool	2023	Covetool, Georgia, US	Atlanta, GA, USA	Black-Box Model		✔			✔	Annual	Pros: Streamlines the process of energy modelling with an emphasis on cost and performance, integrating sustainable design strategies. Cons: As a relatively new entrant, it may not have as wide adoption or comprehensive databases as more established tools.	(Tool)
Edge	3.0	Team of Edge	UK	-	Single-Family House and Multi-Family House	✔				-	Pros: Focuses on sustainability and offers certifications for green buildings, with a user-friendly interface. Cons: Primarily used for certification purposes and may not be as detailed for technical engineering analysis.	(*Edge App*)

Software	Version	Developer	City, Country	Black Box/White Box	System Boundary	Energy	Thermal	Daylight	Air Quality	Time Frame	Pros and Cons	References
DALEC	2023	DALEC Team	-	Black-Box Model	Single-Family House and Multi-Family House				✔	Monthly	Pros: Provides life-cycle carbon and energy analysis, useful for assessing the environmental impact of buildings. Cons: The focus on carbon may mean that energy efficiency measures are not as comprehensively addressed.	(Werner et al., 2017)
MIT Design Advisor	1.1	MIT Department of Architecture	MA, USA	Black-Box Model	Neighborhood and District Scale					Annual	Pros: Allows for quick assessment of design strategies on building energy use, targeted toward the early design phase. Cons: Limited in scope and may not be suitable for detailed final analysis or large-scale projects.	(Urban & Glicksman, 2006)
HeliOS EE-SIM	2017	Helios Inc.	CA, US	Black-Box Model	Neighborhood and District Scale					Monthly	Pros: Specializes in solar potential and energy simulation, aiding in the design of photovoltaic systems. Cons: Focus on solar analysis means other aspects of building energy management might need additional tools.	(Lee et al., 2015)
Neo Net Energy Optimizer	2023	Ryan schwartz	Canada	Black-Box Model	Single-Family House and Multi-Family House	✔					Pros: Specializes in optimizing net-zero energy buildings by balancing energy production and consumption, making it valuable for sustainable design projects. Cons: Focus on net-zero energy optimization may not be as comprehensive for general energy management needs or less sustainable-oriented projects.	(*NEO Net Energy Optimizer*)
SEMERGY	2016	XYLEM Technologies	-	White-Box Model	Multi-Districts and City Scale	✔	✔				Pros: Utilizes a web-based decision support system for optimizing energy efficiency in building renovation, incorporating a broad range of data including climate, building materials, and systems. Cons: May require detailed inputs and specific knowledge of renovation projects, which can limit its utility for initial design stages or new constructions.	(*SEMERGY - Energy Efficient Buildings*)
Energinet	2021	Cebyc AS	Denmark	Black-Box Model	Multi-Districts and City Scale	✔					Pros: Aims to provide a comprehensive database and networking platform for energy market data, potentially facilitating energy trading and market analysis. Cons: Its role as a data platform means it may not directly assist in building-specific energy modelling or management tasks.	(*Energinet Energy Management Software*)
EPWMap	0.0.6	Mostapha Roudsari	-	White-Box Model	Single-Family House and Multi-Family House				Air Quality		Pros: Offers an interactive map of EnergyPlus weather data, aiding in the selection of appropriate climate data for building energy simulations. Cons: As a tool focused on climate data provision, it does not perform energy modelling or analysis itself.	(*EPWMAP*)
Deksoft	2.1	Petr Kocian	Czech Republic	White-Box Model	Single-Family House and Multi-Family House		✔				Pros: May refer to software tools designed for specific energy management tasks, possibly including building performance analysis. Cons: Without more context on Deksoft, it is challenging to provide specific pros and cons; if it is specialized software, it may have limited applicability or require specialized knowledge to use effectively.	(*DEKSoft*)
GEnergy	-	Donald alexander	USA	White-Box Model	Single-Family House and Multi-Family House	✔				Hourly, Sub Hourly	Pros: Can offer user-friendly interfaces for energy auditing and management, aiming to simplify the process of identifying energy-saving opportunities. Cons: May lack the depth of more specialized simulation tools for detailed technical analysis.	(*GEnergy*)

Software	Version	Developer	City, Country	Black Box/White Box	System Boundary	Energy	Thermal	Daylight	Air Quality	Time Frame	Pros and Cons	References
EnExPlan	-	Marc Lacombe Almiranta	Montreal, QC, Canada	White-Box Model	Neighborhood and District Scale	✔					Pros: Designed for energy exploration and planning, this tool may assist in strategizing energy distribution and conservation measures. Cons: It might be more suited for macro-level planning rather than detailed building-specific simulations.	(*EnExPlan*)
ReOpt Lite	3.0.1	Linda Parkhill - NREL	USA	Black-Box Model	Single-Family House and Multi-Family House			✔		Hourly/Annual Analysis	Pros: Provided by the National Renewable Energy Laboratory (NREL), it helps optimize energy systems for cost and performance, focusing on renewable integration and grid reliability. Cons: As a "lite" version, it may not include all the features of a full-scale model, potentially limiting detailed analysis.	(*REopt*)
Hippo CMMS	-	Daniel Golub	Winnipeg, Canada	Black-Box Model	Single-Family House and Multi-Family House	✔					Pros: Offers a computerized maintenance management system (CMMS) that can track and manage building maintenance operations, indirectly affecting energy efficiency through optimal equipment performance. Cons: Its primary focus is on maintenance management rather than direct energy modelling or simulation.	(*Hippo CMMS*)
Building performance database (BPD)	-	Robin Mitchell	USA	-	Multi-Districts and City Scale	✔					Pros: The largest publicly available source of building performance data in the U.S., useful for benchmarking and analyzing building energy performance. Cons: Primarily a database, it does not perform simulations or analyses but requires interpretation of data for application in energy management.	(*Building Performance Database (BPD)*)
Snugg PRo	5.0	Sandy Michelas	Denver, CO, USA	Black-Box Model	Single-Family House and Multi-Family House	✔					Pros: A software tool tailored for home energy audits that can provide recommendations for energy efficiency improvements and detailed reports. Cons: May not be as comprehensive for commercial buildings or large-scale energy management projects, only for US	(*Snuggpro*)

The web tool approach does not necessitate an in-depth comprehension of heat transfer and thermal behaviour in building components, as is the case with physics-based models. Therefore, this approach suits building circumstances where physical characteristics are not determined. Due to this feature, design and retrofit toolkits that use data-driven calculation methods are also ideal for non-expert users, such as designers and building owners (Foucquier et al., 2013).

Due to the capability and application of BES systems, several inputs and outputs for these BES tools are depicted in Figure 2.11.

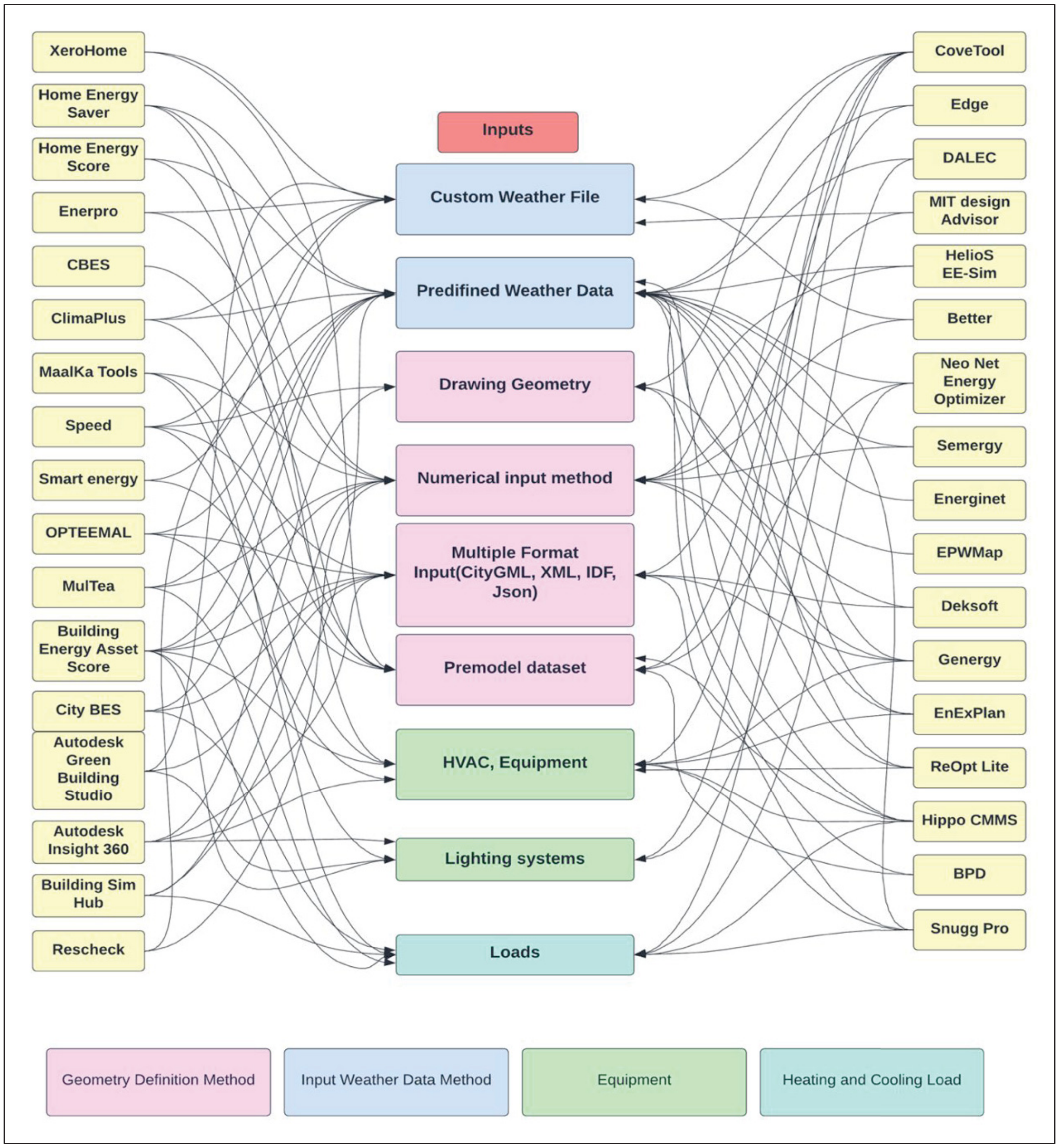

Figure 2.11 Input methods for Webtool Energy Simulation tool

These toolkits have many other capabilities, as listed in (Crawley et al., 2008; Forouzandeh et al., 2021). Fourteen out of the 34 can perform load analyses based on wall, ceiling, and window materials. Building material input is possible with web applications, such as Home Energy Saver and Home Energy Score. Enhancing and performing parametric investigations enhance the utility of toolkits, particularly during the initial stages of design when numerous design alternatives and tactics are accessible, thereby aiding decision-making.

XeroHome and Cove Tool are two of the five web-based toolkits capable of optimizing building energy management. XeroHome and Insight 360 do not implement optimization through parametric simulation, but SPEED, Cove Tool, and BuildSimHub do. These tools present users with the best design options based on all possible combinations of design features. Various toolkits provide diverse modes for defining input data, catering to users with varying proficiency levels and engagement in different phases of building design or retrofit. Examples include XeroHome, Home Energy Saver, Cove Tool, and CBES, which are suitable for users with differing levels of expertise. CoveTool has a wide range of capabilities, including water analysis, energy analysis, carbon emissions analysis, and economic analysis. It uses machine learning algorithms to analyze thousands of alternatives for better analyzing building energy systems.

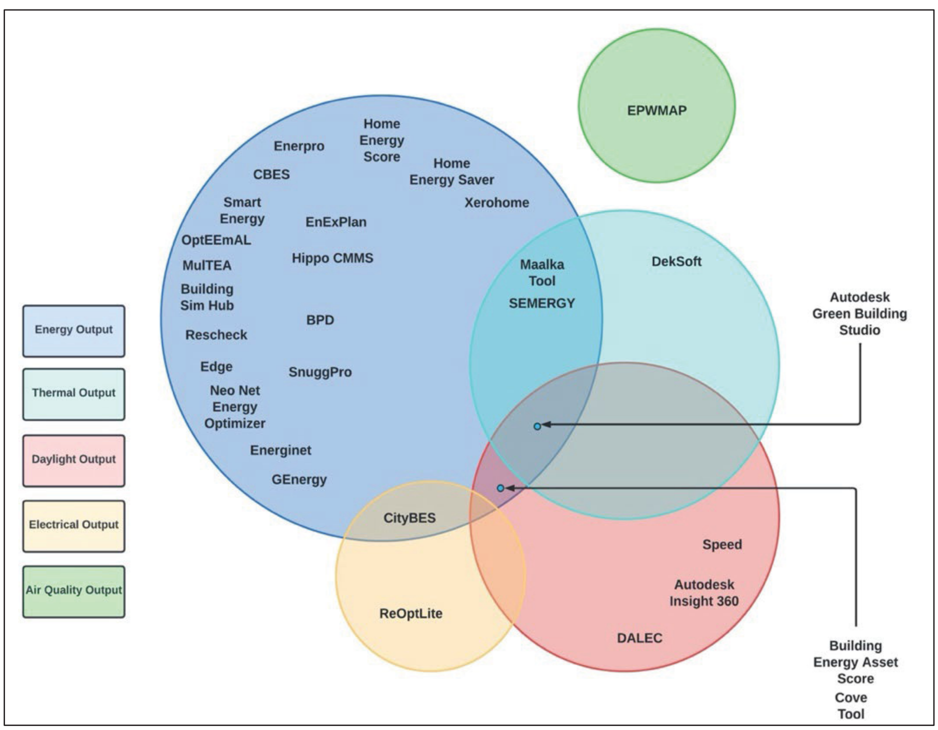

Figure 2.12 Outputs for Webtool Energy Simulation tool

The comparative review of white box and black box simulation tools presented in this chapter is directly applicable to building energy consulting practices. By outlining the strengths, limitations, and use cases of tools such as EnergyPlus, TRNSYS, CitySim, and IDA ICE, the review helps consultants select modeling approaches tailored to specific project requirements, whether for early-stage design analysis, retrofit evaluation, or compliance verification. For example, white box tools are ideal when detailed physical modeling is needed for HVAC design or energy certification, whereas black box and data-driven models (e.g., ANN-based predictors) are more suitable for forecasting, real-time analytics or large-scale urban energy assessments where historical data is available. This framework enables consultants to align tool capabilities with key project constraints such as data availability, computational complexity, client interpretability, and regulatory context. Therefore, the review functions as a decision-support reference, guiding tool selection, and model implementation strategies in practical consulting scenarios.

While all reviewed tools offer valuable capabilities, their suitability depends on specific project requirements. For high-resolution, physically detailed modeling, tools like EnergyPlus or TRNSYS are preferred due to their robustness and flexibility, especially in new construction or compliance-driven projects. However, these tools require substantial data input and expertise. In contrast, tools such as CitySim or black-box ML models are better suited for urban-scale or data-limited contexts, where fast approximation and decision support are more critical than granular thermal dynamics. Web-based tools, though limited in customization, are practical for consultants needing quick results for client communication or early-stage feasibility analysis. Table 2.7 summarizes tool suitability based on criteria such as data availability, modeling resolution, project scale, and regulatory compliance needs.

This comparative framework allows practitioners to align tool selection with the specific objectives, constraints, and timelines of real-world energy consulting projects.

2.6 Conclusions and Recommended Future Research

2.6.1 Conclusions

In conclusion, this comprehensive review of building energy management simulation tools, now enriched with insights into their relevance to consulting practices, underscores the vast array of tools available to address the multifaceted challenges of optimizing energy efficiency in buildings. The tools encompass both white box models, rooted in fundamental physics principles, and black box models, harnessing the power of machine learning and statistical approaches. Including web-based simulation tools further expands the toolbox, providing accessibility and flexibility in data visualization and benchmarking. Consulting practices in the field of building energy management benefit immensely from these diverse tools. White box models offer intricate insights into energy system operation and are invaluable for detailed energy performance analysis, particularly in retrofitting projects and compliance verification. On the other hand, black box models provide consultants with rapid results, making them ideal for operational optimization, predictive maintenance, and real-time energy management.

The choice of modelling approach and tool depends on the specific consulting objectives and building complexities. Large and complex buildings demand scalable tools, emphasizing the importance of selecting the right tool for the desired level of accuracy. This review underscores the significance of a tailored approach, highlighting the plethora of tools available to support data-driven decision-making, optimize energy systems, and enhance energy efficiency in the built environment. Furthermore, integrating these tools into consulting practices enables data scientists, analysts, and engineers to collaborate effectively, providing actionable insights and facilitating informed decision-making. As the consulting industry continues to evolve, building energy management simulation tools remains at the forefront, empowering professionals to address energy challenges, enhance sustainability, and drive innovation in the design and operation of buildings. The dynamic landscape of building energy management demands adaptability and expertise, and this review serves as a valuable resource for navigating this ever-evolving field.

2.6.2 Recommendations for future research

Based on current knowledge, several areas require further research to advance the development and application of building energy simulation tools, specifically white box and black box models. In addition, six knowledge gaps in building energy simulation tools can be addressed in future studies.

1. Lack of standardized methodology: Despite the availability of numerous simulation tools, there needs to be a standardized methodology for comparing and evaluating these tools. Future studies can address this gap by proposing a standardized methodology that can be used for consistent evaluation of simulation tools.

2. Limited studies on the accuracy of black box models: While black box models are gaining popularity in building energy management, there is a limited number of studies on their accuracy compared to white box models. Future studies can address this gap by conducting comprehensive accuracy studies of black box models and comparing them with white box models.

3. Limited studies on the scalability of white box models: While white box models are considered accurate, their scalability to larger building complexes or districts is a concern. Future studies can address this gap by investigating the scalability of white box models and developing methods to improve their scalability.

4. Lack of integration between white box and black box models: White box and black box models are often used separately, and there needs to be more integration between them. Future studies can address this gap by exploring ways to combine both types of models to improve accuracy and scalability.

5. Limited studies on the impact of uncertainties on model predictions: More studies are required on the impact of uncertainties on model predictions, which is crucial for decision-making in building energy management. Subsequent research endeavours have the potential to fill this void by quantifying the influence of uncertainties on model predictions and devising approaches to enhance the resilience of simulation tools.

6. Limited studies on the usability and accessibility of simulation tools: While simulation tools are becoming more advanced, there is a lack of studies on their usability and accessibility, particularly for non-expert users. Future studies can address this gap by evaluating the usability and accessibility of simulation tools and developing user-friendly interfaces for non-expert users.

7. In conclusion, implementing these recommendations in future studies can lead to a more precise evaluation of building simulation tools and help address the knowledge gaps identified in this review.

Funding: This research received no external funding

Conflicts of Interest: The authors declare no conflict of interest

Nomenclature

BEMS	Building Energy Management Systems
BES	Building energy simulation tools
BIM	Building Information modelling
BPS	Building performance simulation tools
DDM	Data-Driven Modelling
DL	Deep Learning
DNN	Deep Neural Networks
HVAC	Heating, ventilation, and air conditioning
IFC	Industry foundation class

LSTM	Long Short-Term Memory
LR	Linear Regression
LTLF	Long-term load forecasting
ML	Machine Learning
NMF	Neutral model format
PCM	Phase Change Material
RF	Random Forest
SVM	Support Vector Machine
STLF	Short-term load forecasting

CHAPTER 3

Advanced Machine Learning Techniques for Energy Consumption Analysis and Optimization at UBC Campus: Correlations with Meteorological Variables

Amir Shahcheraghian [a], Adrian Ilinca [a]

[a] Department of Mechanical Engineering, École de Technologie Supérieure, 1100 Notre-Dame West, Montreal, Quebec, Canada H3C 1K3

Paper Published in *MDPI-Energies*, September 2024

Abstract

Energy consumption analysis has often faced challenges with limited model accuracy and inadequate consideration of the complex interactions between energy usage and meteorological data. This study is presented as a solution to these challenges through a detailed analysis of energy consumption across UBC Campus buildings using a variety of machine learning models, including Neural Networks, Decision Trees, Random Forests, Gradient Boosting, AdaBoost, Linear Regression, Ridge Regression, Lasso Regression, Support Vector Regression, and K-Neighbors. The primary objective is to uncover complex relationships between energy usage and meteorological data, addressing gaps in understanding how these variables impact consumption patterns in different campus buildings by considering factors such as seasons, hours of the day, and weather conditions.

Significant interdependencies among electricity usage, hot water power, gas, and steam volume are revealed, highlighting the need for integrated energy management strategies. Strong negative correlations between Vancouver's temperature and energy consumption metrics are identified, suggesting opportunities for energy savings through temperature responsive strategies, especially during warmer periods. Among the regression models evaluated, deep neural networks are found to excel in capturing complex patterns and achieving high predictive accuracy.

Valuable insights for improving energy efficiency and sustainability practices are offered, aiding informed decision-making for energy resource management on educational campuses and similar urban environments. The application of advanced machine learning techniques underscores the potential for data-driven energy optimization strategies. Future research could investigate causal relationships between energy consumption and external factors, assess the impact of specific operational interventions, and explore integrating renewable energy sources into the campus energy mix. Through these efforts, UBC can advance sustainable energy management and serve as a model for other institutions aiming to reduce their environmental impact.

Keywords: energy consumption, machine learning, meteorological data, regression models, energy efficiency, UBC Campus, Neural Networks, electricity usage

3.1 Introduction

In 2018, global building emissions increased by 2% for the second consecutive year, reaching 9.7 gigatons of carbon dioxide ($GtCO_2$) (Hamilton et al., 2020). This marks a shift from the trend observed between 2013 and 2016, when emissions stabilized. The floor space and population growth led to a 1% rise in energy consumption in buildings, reaching 125 exajoules (EJ), which constitutes 36% of global energy use (Hamilton et al., 2020). Academic buildings on a typical UK university campus cover about 42% of the total area and are responsible for nearly 50% of the campus's energy consumption and carbon emissions (Hung, 2020). Predicting energy usage in these buildings has become increasingly important for facility managers, aiding energy planning and regulation (Hung, 2020). Accurate forecasting can reveal potential energy savings and support energy conservation efforts on campus (Han et al., 2022). However, energy consumption is influenced by various factors, including meteorological conditions, electromechanical system operations, and holiday schedules, making predictions challenging due to the inherent nonlinearity and uncertainty (Han et al., 2022).

There are two main approaches to predicting building energy consumption: physical and data-driven models (Ding & Liu, 2020). Physical models analyze buildings based on physical principles but require extensive data and long simulation times, making them complex and less efficient. In contrast, data-driven models use historical data to predict energy consumption and are known for their flexibility, ease of data acquisition, and accuracy (Ding & Liu, 2020).

This study evaluates different data-driven models, including machine learning and deep learning techniques, to assess their effectiveness in predicting energy consumption for campus buildings. Specifically, it compares various machine learning models (e.g., Decision Tree, Random Forest, Gradient Boosting) and neural networks (ANN) across 167 buildings on the UBC Campus. The analysis examines prediction accuracy using metrics such as MAE and R^2 scores, providing insights into the performance of these models.

The goal of this article is to conduct a detailed analysis of energy consumption within UBC Campus buildings using a range of machine learning models, including Neural Network, Decision Tree, Random Forest, Gradient Boosting, AdaBoost, Linear Regression, Ridge Regression, Lasso Regression, Support Vector Regression, and K-Neighbors. The study explores how different factors such as seasons, time of day, and meteorological conditions affect energy consumption patterns. The research uses sophisticated analytical methods to uncover and quantify the complex relationships between energy demand and environmental factors. The findings offer valuable insights for improving energy efficiency, sustainability practices, and informed decision-making for energy management in educational campuses and similar urban settings. A key knowledge gap in energy consumption analysis is the limited accuracy of traditional models that fail to capture the complex, nonlinear interactions between energy usage and meteorological data, resulting in suboptimal predictions and reduced energy management effectiveness. Addressing this gap requires advanced models that better understand these dynamics, ultimately improving accuracy and energy optimization efforts.

To address the goal of conducting a detailed analysis of energy consumption within UBC Campus buildings using a variety of machine learning models, it is crucial to examine the challenges faced in previous studies and how they highlight the need for more advanced techniques. One of the primary issues identified in earlier research is the reliance on relatively simple statistical models, which, while easier to implement, often fall short of capturing the full complexity of energy consumption dynamics. For instance, Khuram Pervez Amber et al . (Amber et al., 2017) developed a forecasting model using Multiple Regression (MR) to predict daily electricity consumption in university buildings based on six explanatory variables, including ambient temperature, solar radiation, and building type. Although the model achieved promising results, with a Normalized Root Mean Square Error (NRMSE) of 12% and 13% for two different building types, it struggled to account for deeper non-linear interactions between these variables. The study emphasized the need for more reliable and flexible models capable of accommodating a broader range of buildings and capturing more complex relationships between variables.

Similarly, Sadeghian Broujeny et al. (Sadeghian Broujeny et al., 2023) outlined the challenges in forecasting building energy consumption due to the intricate management of multiple influencing parameters. Their work demonstrated the effectiveness of artificial intelligence (AI) models, such as Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU), achieving an RMSE of 0.23. However, they also highlighted the need for selecting optimal time lags and exogenous data to enhance model performance. While these AI models provided superior predictive accuracy, challenges remain in managing model complexity and extracting meaningful historical knowledge from a wide range of features.

These studies underscore the knowledge gap in developing models that can accurately forecast energy consumption while capturing the complex, nonlinear interactions between meteorological variables and energy usage. This article aims to bridge this gap by employing a range of machine learning models, including Neural Network, Decision Tree, Random Forest, Gradient Boosting, AdaBoost, Linear Regression, Ridge Regression, Lasso Regression, Support Vector Regression, and K-Neighbors, to explore these interdependencies more effectively.

By doing so, this study sets out to establish new baseline accuracy levels and address the limitations of prior work, where simpler models or narrowly focused AI techniques may have missed critical patterns in energy consumption. The application of these advanced models will provide a more comprehensive understanding of how factors such as temperature, time of day, and seasonal variation influence energy use in campus environments, establishing a new benchmark for future research.

3.2 Literature Review and Background

Lei et al. (Lei et al., 2021) developed a prediction model for building energy consumption by integrating rough sets with deep learning, utilizing a data-driven and adaptive control scheme. Their model combines rough set theory, which uses a genetic algorithm-based attribute reduction method to eliminate redundant factors, with a deep belief network (DBN) for information recognition. This hybrid approach significantly improves prediction accuracy for both short term and medium term energy consumption compared to traditional neural networks like BP, Elman, and fuzzy neural networks. The use of rough set theory reduced the number of inputs, thereby enhancing the performance of the DBN. This advancement highlights the potential of machine learning in creating accurate building energy simulations, enabling more effective real-time power supply scheduling and demand management.

Fan et al. (Fan et al., 2017) explored the effectiveness of deep learning techniques for predicting building cooling loads, comparing them with state-of-the-art prediction methods and feature extraction techniques. They found that deep learning methods, particularly nonlinear techniques like extreme gradient boosting (XGB), outperformed linear methods. The best performance was achieved using XGB models with features extracted by unsupervised deep learning models, such as deep autoencoders. Interestingly, supervised deep learning models did not require complex architectures; a shallow model with two hidden layers sufficed, likely due to dataset size and inherent patterns.

This study underscores the value of unsupervised deep learning in feature extraction, which consistently improves prediction accuracy over conventional methods. The proposed deep learning-based techniques offer precise 24-hour-ahead cooling load predictions, which help build operation management, demand side management, optimal control strategies, and fault detection. Future research should explore diverse data sources to validate and expand these methods.

Jui Sheng Chou et al. (Chou & Tran, 2018) reviewed machine learning techniques for energy consumption forecasting in buildings using real-time smart grid data. They highlighted the effectiveness of hybrid models that combine forecasting and optimization techniques, demonstrating improved accuracy and usability for energy management planning. This review contributes to advancing energy efficiency and sustainable development efforts.

Carla Sahori Seefoo Jarquin et al. (Seefoo Jarquin et al., 2023) examined five university buildings to create day ahead load forecasts using feed-forward neural networks trained with a similar day method. They classified data into low, medium, and high consumption categories and evaluated ten models per category against benchmarks. Most models exceeded benchmarks and met ASHRAE accuracy standards, with notable improvements from incorporating features like seasonal shifts, day-night cycles, and recency effects. Models M3, M9, and M10 were remarkably accurate. Despite robust point forecasts, the prediction intervals from a dynamic ensemble approach were too narrow, suggesting the need for alternative methods. This research indicates the potential for extending forecasts to heat demand, various facility types, smart cities, and energy communities, emphasizing hierarchical forecasting for better collective energy management.

Bo Han et al. (Han et al., 2022) analyzed the use of support vector machine (SVM), Gaussian process regression (GPR), and decision tree (DT) models for predicting campus building energy consumption.

Their comparative study found that the SVM model achieved the highest prediction accuracy, outperforming the GPR model by 5.79% and the DT model by 13.93%. This study demonstrates the feasibility and effectiveness of the SVM model in accurately predicting electricity consumption in campus buildings, highlighting its superiority over other prediction methods.

Bilal Akbar et al. (Hung, 2020) investigated two data-driven forecasting techniques, artificial neural networks (ANN) and multiple regression (MR), to estimate daily electricity usage at London South Bank University. Using historical data from 2007 to 2011, they analyzed the impact of five different climate factors on energy consumption. They found that energy consumption is highly influenced by the Weekday Index, a proxy for building occupancy, and Dry Bulb Temperature. ANN outperformed MR, achieving mean absolute percentage error (MAPE) values of 2.44% for working days and 4.59% for non-working days. Both techniques performed well, but ANN demonstrated a slight advantage. This research provides a foundation for predictive studies in other building categories and offers valuable insights for energy managers seeking to accurately forecast building energy usage patterns. The findings suggest that predictions remain reliable over time if the building operates under the same schedule.

Linas Gelažanskas et al. (Gelažanskas & Gamage, 2015) analyzed hot water power data from residential buildings to develop and test various forecasting models, including exponential smoothing, SARIMA, and seasonal decomposition. The results show that these models outperform simpler benchmarks, with seasonal decomposition being the most effective in improving forecast accuracy. This work highlights the significance of advanced predictive techniques for enhancing demand-side management and supporting grid stability in the context of increasing renewable energy integration. Jinyuan Liu et al. (Liu et al., 2021) reviews 70 years of gas consumption forecasting, categorizing the evolution into four stages: initial, conventional, AI, and all round. They highlight that time series models excel in long-term forecasting with a mean absolute percentage error of 1.90%, and support vector regression models (4.98%) are more suitable for short-term forecasting.

The review emphasizes the impact of advancements in computer science and AI on forecasting performance and proposes a framework for model selection along with future research directions. Predicting gas volume, electricity, and hot water power consumption on campuses poses several challenges that Machine Learning (ML) and Artificial Neural Networks (ANN) can address effectively. For gas volume consumption, issues include the impact of seasonality and weather, varying usage patterns across buildings, and data sparsity. ML and ANN models can tackle these challenges by integrating weather data and using regression techniques to correlate consumption with relevant features. Electricity consumption prediction is complicated by complex load patterns, demand fluctuations, and the need to account for multiple factors such as HVAC systems and lighting. Advanced ML models, including ANN, can enhance prediction accuracy. Similarly, predicting hot water power consumption involves addressing challenges related to temperature variability, usage patterns, and system dynamics. Employing predictive models, making seasonal adjustments, and incorporating system performance data can improve accuracy. Overall, effective data preprocessing, feature selection, and model evaluation are crucial for developing robust and accurate consumption predictions.

3.3 Research Methodology

3.3.1 Data Collection

The data collection for this study utilizes the UBC Energy & Water Services' online live data, accessible through the SkySpark platform (U. University). This platform is part of the Campus as a Living Lab initiative and offers valuable insights into building energy and performance metrics. SkySpark, developed by Skyfoundry, is an Internet of Things (IoT) platform that leverages heating, ventilation, air conditioning (HVAC), and energy data to enhance building performance (U. University).

Since 2017, the Energy & Water Services' Energy Conservation team has used SkySpark to monitor and identify energy and greenhouse gas (GHG) savings opportunities. The platform supports long-term storage for a substantial amount of data, including 68,000 data streams, 370 energy meters across 167 campus buildings, 5,000 pieces of HVAC equipment, and 3 Building Management Systems (BMS) vendors (U. University).

For this study, data from 2023 was collected, encompassing hourly profiles from January 1 to December 31. This comprehensive dataset provides a detailed view of energy consumption and performance metrics across various building types and functions on the UBC campus, as illustrated in Figure 3.1.

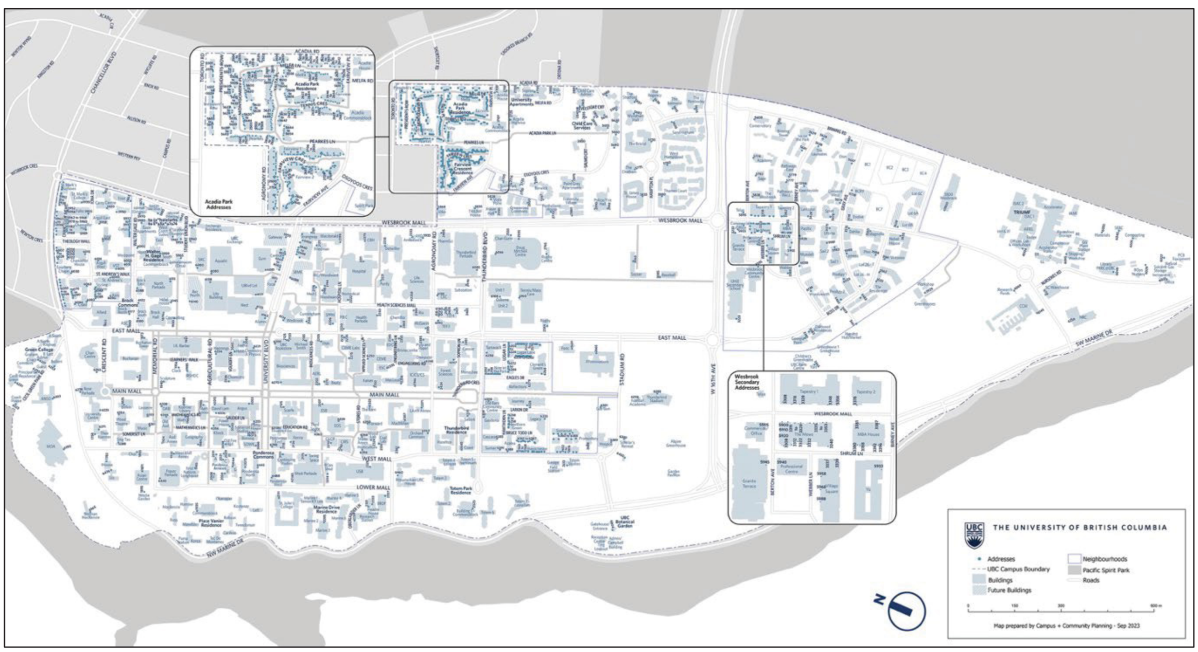

Figure 3.1 UBC campus building map

The units used throughout this study reflect the conventions of the source utility data and underlying measurement systems. Hot water power consumption is expressed in kilowatts (kW) to represent the instantaneous thermal energy delivery rate from district systems. Natural gas use is recorded in cubic meters (m^3), which corresponds to volumetric fuel consumption used by central boilers; this aligns with how utilities report and bill gas usage. Steam consumption is expressed in pounds (lb), following North American conventions where steam systems are typically monitored by mass due to variable volumetric properties under changing temperature and pressure. While these units differ in type (power, volume, mass), each aligns with standard engineering practice and the measurement infrastructure in place at the UBC campus.

3.3.2 Data Pre-Processing

The initial phase of the study involves collecting a diverse range of data frames crucial for analyzing energy consumption:

- Time-Dependent Data Frames: Capturing temporal patterns in energy usage.
- Non-Time-Dependent Data Frames: Providing static information about building characteristics.
- Weather Data Frames: Offering environmental context, including meteorological conditions.
- Electricity Energy Data Frames: Detailing power usage across the campus.
- Gas Volume Data Frames: Indicating gas consumption patterns.
- Hot Water Energy Data Frames: Tracking hot water usage.
- Steam Volume Data Frames: Recording steam consumption.
- Water Volume Data Frames: Monitoring general water usage.
- Total Data Frames: Consolidating comprehensive building energy data for each building in the target year, 2023.

This extensive dataset establishes the foundation for the subsequent analysis.

The steam system at UBC is primarily used for space heating and domestic hot water generation in older buildings across campus. High-pressure steam is generated at the central energy plant and distributed through an underground network of insulated pipes. In buildings served by this system, steam is converted via heat exchangers to heat water or air. The system plays a significant role during colder months and exhibits strong correlation with ambient temperature.

The hot water system refers to low-temperature hot water (LTHW) used for space heating purposes, particularly in newer or retrofitted buildings on campus. It is circulated through a hydronic network and used in radiant heating systems, baseboards, or HVAC coils. This hot water is typically produced using natural gas-fired boilers or steam-to-hot-water heat exchangers. As it is directly linked to building thermal load, its energy use shows strong seasonal variation and correlates with temperature and time-of-year indicators.

Natural gas is used primarily as a fuel for steam and hot water generation at UBC. The majority of campus boilers are gas-fired, and gas consumption is therefore closely tied to both steam and hot water demand. In some cases, gas may also support standalone HVAC units or emergency heating systems. Because gas is upstream of both thermal systems, its use often correlates with both hot water and steam energy consumption, especially during heating seasons.

Following data collection, rigorous data cleaning procedures ensure data integrity and reliability. Figure 3.2 illustrates the total electricity energy and hot water power consumption across the campus for 2023. Outliers for Electrical Energy and Hot Water Power are identified and removed from crucial data frames to prevent skewed results and enhance predictive accuracy. Data quality checks are also performed on weather and electrical energy data frames to eliminate any NaN (Not a Number) values, ensuring the dataset is complete and reliable.

The decision not to use pipelines and transformers for preprocessing was based on the specific requirements of this study, which focused primarily on energy consumption forecasting. While pipelines and transformers can streamline preprocessing tasks and ensure a consistent workflow, this study employed a more targeted approach for data cleaning and outlier removal, using tailored procedures such as the Interquartile Range (IQR) method to ensure the reliability of the dataset.

Outliers were estimated using the IQR method, a robust statistical approach commonly applied to detect and remove extreme values in the dataset. The IQR method calculates the difference between the third quartile (Q3) and the first quartile (Q1), identifying outliers as any data points lying below Q1 - 1.5 * IQR or above Q3 + 1.5 * IQR. This method was chosen for its simplicity and effectiveness in dealing with energy consumption data, where occasional extreme values can significantly distort model performance.

Before outlier deletion, the dataset comprised 8,760 rows, representing hourly data for the entire year. After applying the IQR method, 31 outliers were removed, resulting in a final dataset of 8,729 rows. This minimal reduction in data ensures that the dataset remains comprehensive while improving the overall quality and accuracy of the subsequent machine learning models. By focusing on high-quality data, the models developed in this study can more accurately capture energy consumption patterns and their dependencies on meteorological variables.

A total values column is added to each data frame to view energy consumption comprehensively. This column aggregates the energy consumption for each hour by summing the values across all buildings, providing a consolidated overview of energy usage patterns. Following preprocessing, a range of machine learning models, including Neural Network, Decision Tree, Random Forest, Gradient Boosting, AdaBoost, Linear Regression, Ridge Regression, Lasso Regression, Support Vector Regression, and K-Neighbors, are applied to the cleaned dataset.

The selection of machine learning models for this study was driven by the need to address the complex and nonlinear relationships between energy consumption and meteorological variables, as well as to compare a range of model types to determine the most effective approach. Each model was chosen based on its strengths in handling different aspects of the data and the challenges of energy consumption forecasting.

The Neural Network (ANN) was selected for its ability to model intricate, non-linear relationships between energy consumption and environmental factors like temperature and wind speed. ANN's flexibility, especially with multiple layers and the use of non-linear activation functions such as tanh, makes it particularly effective for datasets with both positive and negative inputs, capturing complex patterns in the data. Decision Trees were chosen for their interpretability and capability to manage non-linear interactions between features, making them useful for understanding how variables such as time of day and seasonal changes impact energy usage.

Building on Decision Trees, Random Forest was included for its ability to reduce overfitting by averaging predictions across multiple trees, resulting in more accurate and stable forecasts across buildings with varying consumption patterns. Gradient Boosting and AdaBoost, both ensemble methods, were chosen for their superior performance through the iterative improvement of weak learners. Gradient Boosting is particularly well-suited for regression tasks, providing high accuracy in predicting energy consumption, while AdaBoost focuses on difficult to predict instances, helping to refine predictions.

To compare against more complex models, Linear Regression, Ridge Regression, and Lasso Regression were selected as simpler models. Linear Regression serves as a baseline, while Ridge and Lasso offer regularization techniques to prevent overfitting by penalizing large coefficients, providing insights into how well linear assumptions hold in this context. Support Vector Regression (SVR) was chosen for its ability to handle both linear and non-linear relationships by applying the kernel trick, making it suitable for capturing the dependencies between energy usage and meteorological factors in high-dimensional spaces. SVR is also robust to outliers and generalizes well in smaller datasets.

Lastly, K-Neighbors Regression was selected as a non parametric model that predicts energy consumption based on the behaviour of nearby data points, offering a simple yet effective method for capturing local patterns in energy usage. By employing this diverse set of machine learning models, the study seeks to identify the most effective approaches for accurately forecasting energy consumption while accounting for the complex interactions between variables, ultimately providing a more comprehensive understanding of energy patterns in campus environments.

In the training process, the database consists of 8,760 rows, representing hourly data for the entire year of 2023. The measurements were recorded on an hourly basis throughout this period. The dataset was initially cleaned and preprocessed and then divided into three subsets: 70% for training, 15% for validation, and 15% for testing. While the dataset contains 8,760 hourly records for the full year of 2023, it is acknowledged that the operational schedules embedded in the data may exhibit limited dynamic variation, particularly in relation to behavioral shifts such as holidays, weekends, or exceptional climate events. In this study, the modeling approach primarily focuses on identifying macro-scale consumption trends and forecasting average load behavior across the campus. For future work, incorporating more granular and dynamic schedule inputs, including HVAC operation profiles, occupancy patterns, and calendar effects, could enhance the responsiveness of the models and their applicability in real-time operational control or demand-side management strategies.

The training set was used to fit the models, allowing each algorithm Neural Network, Decision Tree, Random Forest, Gradient Boosting, AdaBoost, Linear Regression, Ridge Regression, Lasso Regression, Support Vector Regression, and K-Neighbors to learn patterns and relationships within the data by adjusting their internal parameters to minimize training error. During training, a portion of the training data was set aside as a validation dataset to finetune hyperparameters and prevent overfitting. This validation process involved assessing model performance on this subset and making adjustments to improve generalization. After training and validation, the models were evaluated on the 15% testing set, which remained unseen during the training and validation phases. This evaluation assessed the models' ability to generalize to new, unseen data, with performance metrics such as MAE, MSE, and R^2 being calculated to determine their accuracy and effectiveness. This method ensures that the models' predictive performance is realistically measured against data they have not been exposed to during training and validation.

All hyperparameters utilized for ML models in this project are detailed in Table 3.1.

Table 3.1 HyperParameters for ML Models

Model	**Best Parameters**
Decision Tree Regressor	max_depth=20, min_samples_split=5, min_samples_leaf=2, max_features=sqrt
Random Forest Regressor	n_estimators=100, max_depth=20, min_samples_split=2, min_samples_leaf=4, max_features=auto
Gradient Boosting Regressor	n_estimators=100, learning_rate=0.1, max_depth=5, subsample=0.9
AdaBoost Regressor	n_estimators=100, learning_rate=0.1, loss=linear
Linear Regression	fit_intercept=True, normalize=False
Ridge Regression	alpha=1.0, solver=auto
Lasso Regression	
Support Vector Regression	C=1.0, epsilon=0.1, kernel=rbf, degree=3
K-Neighbors Regressor	n_neighbors=5, weights=uniform, p=2, algorithm=auto

Special attention is given to the Artificial Neural Network (ANN) hyperparameters, significantly impacting model accuracy. The ANN used in this study features multiple layers: an initial Dense layer with 128 neurons and a tanh activation function, followed by layers with 64, 64, and 32 neurons, respectively, all utilizing the tanh function. While the ANN architecture described featuring an input Dense layer with 128 neurons followed by layers of 64, 64, and 32 neurons with tanh activations was used as a baseline configuration, its performance was evaluated independently for each target variable, including electrical energy, hot water energy, gas volume, and steam volume. Due to varying correlations between the input features (e.g., temperature, wind speed, month) and the different energy types, the architecture was either retained or modified based on validation performance. In cases where performance degradation was observed, adjustments to the number of layers, neuron counts, or learning rates were made, ensuring that the ANN was appropriately tuned to the complexity and nonlinearity of each specific prediction task. The output layer consists of a single neuron for regression purposes. The 'tanh' activation function was chosen for its ability to introduce non-linearity into the model, which is crucial for capturing complex paterns in energy consumption data. It outputs values between -1 and 1, centring the data and helping to address the vanishing gradient problem. Additionally, 'tanh' handles negative inputs effectively, which aligns with the characteristics of our dataset. This choice helps in improving the model's learning dynamics and performance. All hyperparameters utilized in this project are detailed in Table 3.2.

In this study, "hot water" refers to heating hot water, not domestic use. This classification is consistent with the data collected from UBC campus utility meters and the context of Figure 3.2, which presents thermal demand profiles for space and water heating after data cleaning.

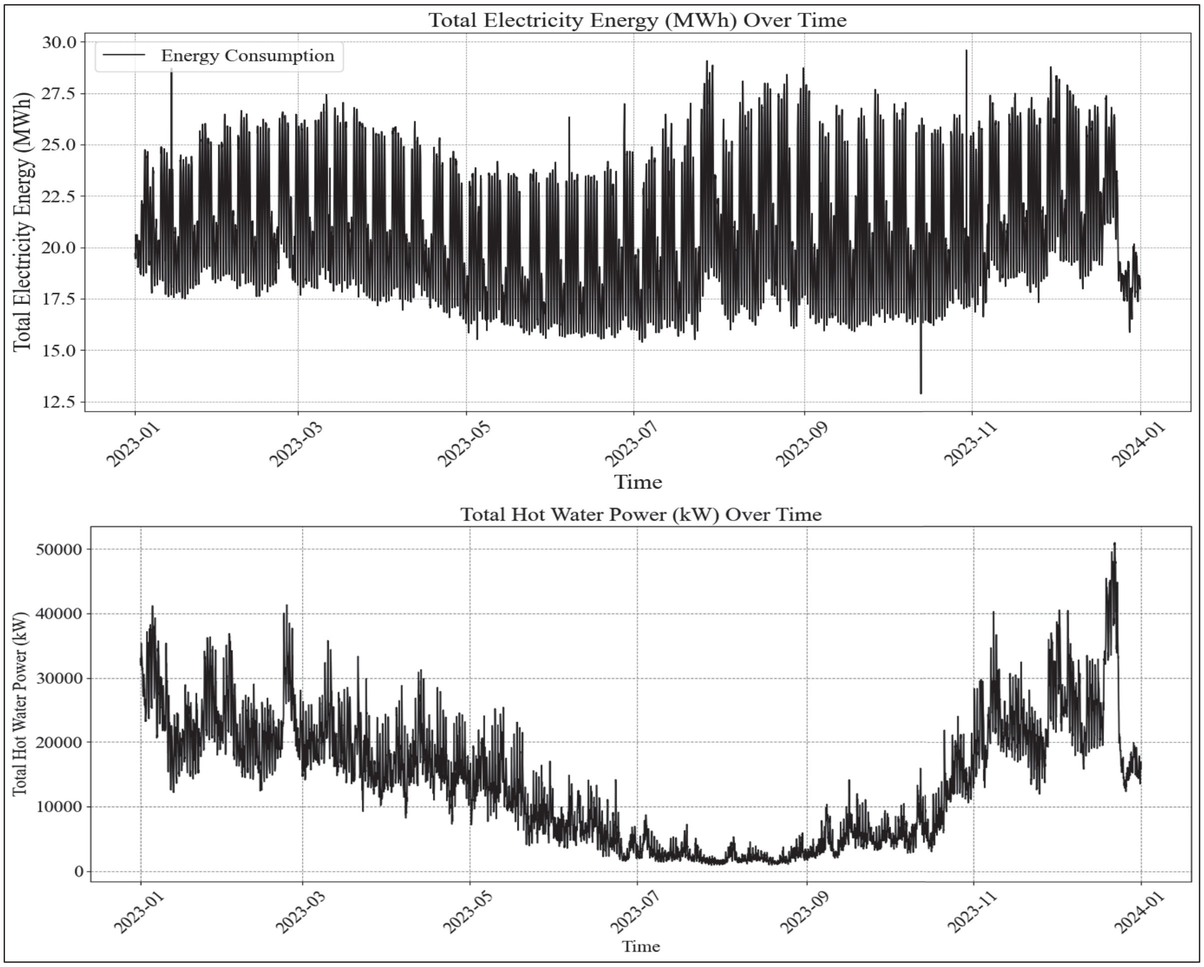

Figure 3.2 Total electricity energy, total heating hot waterpower of UBC Campus in 2023

Table 3.2 Hyperparameters of the ANN Model

Category	**Hyperparameter**	**Value/Description**
Layer Structure	First Layer	Dense layer with 128 neurons, 'tanh' activation function
	Second Layer	Dense layer with 64 neurons, 'tanh' activation function
	Third Layer	Dense layer with 64 neurons, 'tanh' activation function
	Fourth Layer	Dense layer with 32 neurons, 'tanh' activation function

Category	Hyperparameter	Value/Description
	Output Layer	Dense layer with 1 neuron for regression output
Activation Function	Activation Function	'tanh' (used in all hidden layers) The 'tanh' activation function was chosen for its ability to introduce non-linearity into the model, which is crucial for capturing complex patterns in energy consumption data. It outputs values between -1 and 1, centering the data and helping to address the vanishing gradient problem. Additionally, 'tanh' handles negative inputs effectively, which aligns with the characteristics of our dataset. This choice helps in improving the model's learning dynamics and performance.
Regularization	Dropout Rate	0.4 (applied after each BatchNormalization layer)
	Kernel Regularizer	L2 regularization with factor 0.01 (applied to the weights of the second, third, and fourth Dense layers)
Optimization	Optimizer	Nadam
	Learning Rate	0.0005
	Learning Rate Scheduler	ReduceLROnPlateau (reduces learning rate by 0.5 if validation loss does not improve for 10 epochs, minimum learning rate set to 1e-6)
Training Configuration	Batch Size	16
	Epochs	100
Callbacks	Custom Callback	TestLossCallback (tracks test loss at the end of each epoch)

A dropout rate of 0.4 was implemented following each BatchNormalization layer to mitigate overfitting. This technique randomly sets a fraction of input units to 0 during training updates, enhancing model robustness. Additionally, L2 regularization with a factor of 0.01 was applied to the weights of the second, third, and fourth Dense layers. This penalizes large weights, further reducing the risk of overfitting.

The Nadam optimizer was utilized, combining the advantages of Adam and Nesterov momentum with a learning rate of 0.0005 to optimize the model's performance. To fine-tune the learning rate, the ReduceLROnPlateau callback was employed. This callback reduces the learning rate by half if the validation loss does not improve for 10 epochs, with a minimum learning rate set to 10.

A batch size of 16 was chosen to provide a regularizing effect and often improve generalization. The model was trained for 10 epochs, balancing the risk of overfitting and underfitting. A custom callback, TestLossCallback, was used to monitor the test loss at the end of each epoch, offering additional insights into the model's performance on the test set. Through careful selection and tuning of these hyperparameters, the ANN was optimized for better performance and generalization, with each hyperparameter playing a crucial role in enhancing the model's predictive capabilities. Applying machine learning models and meticulous tuning of the ANN's hyperparameters were vital for capturing complex relationships within the dataset, thus providing valuable insights for optimizing energy systems.

Additionally, data preparation included Decomposition Analysis to uncover underlying trends, seasonality, and residual patterns in the energy consumption data. Decomposition involves breaking the data into three components: trend, seasonality, and residuals (Brownlee, 2020). The trend component shows long-term movements in energy consumption, such as increases or decreases. Seasonality captures recurring patterns, such as daily or monthly fluctuations due to weather or operational schedules. Residuals represent random fluctuations or noise not explained by the trend or seasonality (Venujkvenk, 2023).

The ANN structure that explained is a representative architecture optimized for predicting electrical energy consumption. Other ANN models were customized for different target variables (e.g., hot water, gas), with variations in input features, number of hidden layers, and activation functions depending on performance in validation.

Figure 3.3 illustrates electricity energy's trend, seasonal, and residual components. Analyzing these residuals helps identify unpredictable elements in energy consumption and informs strategic decision-making, including peak consumption periods, resource optimization, and accurate future energy demand predictions.

In this study, meteorological parameters (e.g., temperature) were included in all ANN models due to their strong influence on energy consumption patterns. However, ANNs can be applied without such inputs if the use case or available data warrants it. For this specific work, weather variables significantly improved prediction accuracy and were thus systematically used.

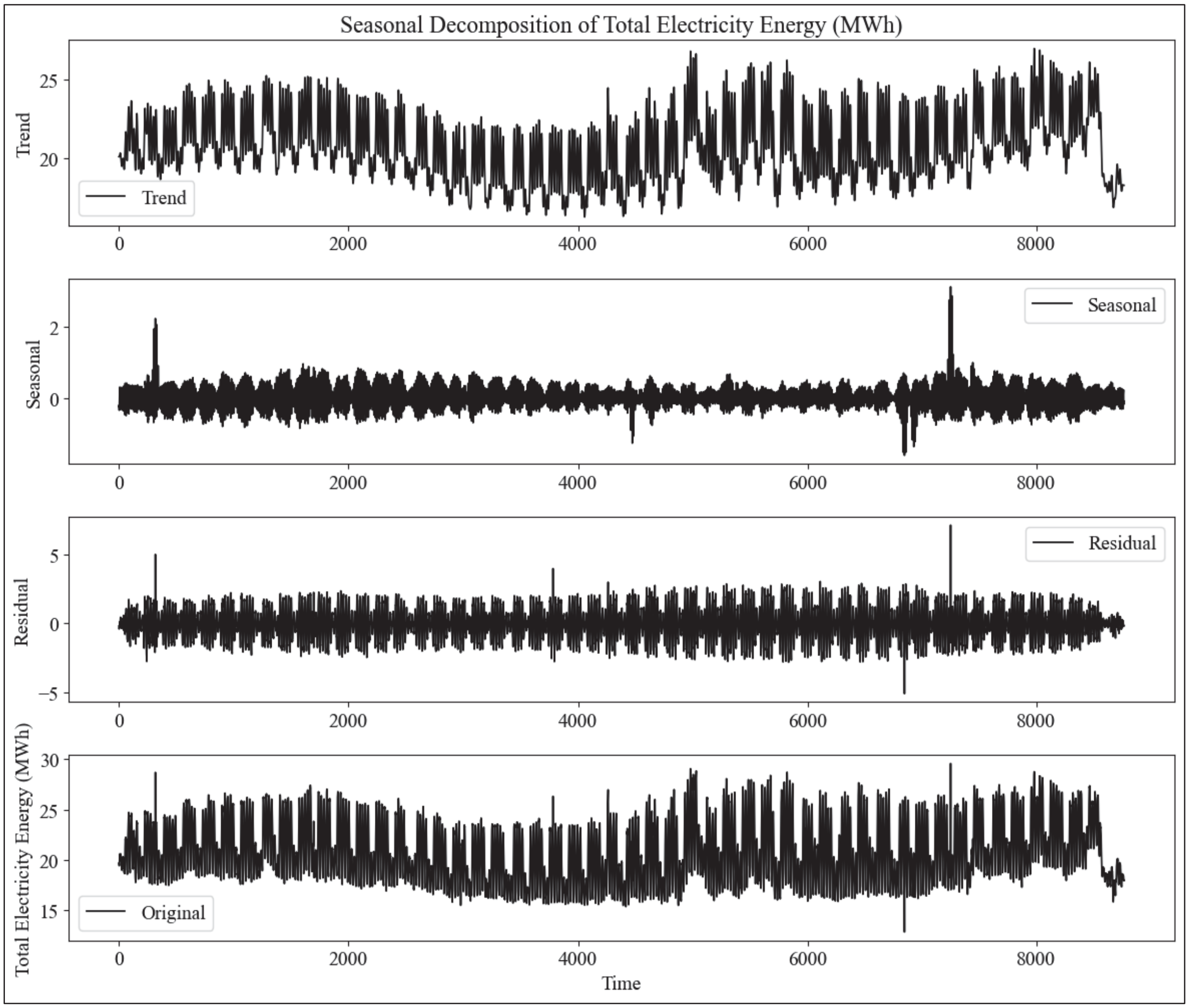

Figure 3.3 Seasonal Decomposition of Total Electrical Energy of 2023

3.3.3 Feature Set and Temporal Resolution

The models for predicting electrical energy, hot water power, and gas volume consumption were developed using a set of key variables obtained from the UBC campus building management system and meteorological data. The input features include ambient temperature, relative humidity, wind speed, solar radiation, and dew point, as well as time-based variables such as hour of the day, day of the week, and seasonal indicators. These variables were selected based on their demonstrated relevance in prior studies of building energy consumption patterns.

The use of hourly time steps was determined by the resolution of the available utility and weather data. While this granularity is sufficient for energy demand forecasting and trend analysis, it is acknowledged that sub-hourly (e.g., 15-minute or 5-minute) intervals may be more suitable for real-time building control and fine-grained operational optimization. Nonetheless, the selected resolution balances model complexity, computational feasibility, and data availability, and it remains appropriate for strategic decision-making in energy planning, load shifting, and system-wide optimization.

3.4 Model Testing and Evaluation

The methodology involves extensive testing and evaluation of various forecasting models to assess their performance in predicting energy consumption. This includes applying different regression methods, such as decision tree regression, to test models across multiple years. This approach allows for a thorough evaluation of predictive accuracy over different time frames.

To determine the effectiveness and reliability of each model, several performance metrics are used:

R-squared (R^2): This statistic measures the proportion of variance in the dependent variable (energy consumption) explained by the independent variables (features) in the model. It ranges from 0 to 1, with higher values indicating a better fit between the model and the data. A higher R-squared value means that the model accounts for a more significant proportion of the variance in the target variable, reflecting better explanatory power. Despite its benefits, R^2 should be interpreted with caution and in conjunction with other metrics to avoid overestimating model performance. Similar studies, including (Sadeghian Broujeny et al., 2023), have employed R^2 to understand the model's ability to capture variability in energy consumption.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{3.1}$$

y_i is the actual value of the dependent variable.
$\hat{y}_i$ is the predicted value of the dependent variable.
$\bar{y}$ is the mean of the actual values of the dependent variable.
n is the number of observations.

Mean Absolute Error (MAE): MAE quantifies the average magnitude of errors between actual and predicted values, providing a clear indication of model accuracy. For instance, similar studies by (Han et al., 2022) and (Wu et al., 2023) have utilized MAE to benchmark model performance in energy forecasting tasks. Lower MAE values signify more accurate predictions, indicating that the model's predictions are closer to the actual values. The advantage of MAE is its simplicity and ease of interpretation; however, it does not penalize larger errors more than smaller ones.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{3.2}$$

Root Mean Square Error (RMSE): RMSE measures the square root of the average squared differences between actual and predicted values. This metric emphasizes more significant errors than MAE, as it penalizes larger discrepancies more heavily. Lower RMSE values indicate better model performance, with more minor deviations from actual values. Its application in previous works (Dong et al., 2024) underscores its effectiveness in evaluating model accuracy, though it shares the disadvantage of sensitivity to outliers.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{3.3}$$

By analyzing these metrics, the methodology evaluates each model's efficacy in capturing energy consumption patterns and provides insights into their predictive accuracy.

While RMSE was initially listed among the evaluation metrics, it was not used in the final performance comparisons for Tables 3.4 to 3.6, which focused on MAE and R^2. This has been clarified to avoid any confusion.

Additionally, we acknowledge that the performance of the models, particularly those for predicting energy use, is context-dependent, as UBC's energy profile includes specific configurations of electricity, gas, and hot water systems. Campuses with different mixes, such as steam networks or higher integration of renewable energy, may yield different patterns.

However, the modeling framework and methodology remain transferable. Other campuses can replicate the approach by using localized data and retraining the models accordingly to capture their specific energy dynamics.

3.5 Results

3.5.1 Correlation between the parameters

The correlation analysis depicted in Figure 3.4 for the UBC University campus in Vancouver reveals several significant relationships among various energy consumption metrics. Notably, the correlation between electricity energy usage and hot water power consumption is 0.38, indicating a moderate positive association where increases in electricity use are moderately related to increases in hot water power consumption. In contrast, the correlation between electricity energy and gas volume is weaker at 0.25, suggesting that changes in one may not strongly impact the other. There is a moderate positive correlation of 0.31 between electricity energy and steam volume, showing a notable relationship between these two metrics.

The analysis also highlights a strong positive correlation of 0.75 between hot water power and steam volume, likely attributed to the implementation of UBC's medium temperature hot water system in 2017 (University, 2024). Furthermore, the correlation between hot water power and gas volume is 0.34, indicating a moderate positive relationship. Gas volume and steam volume exhibit a moderate positive correlation of 0.49, suggesting a considerable relationship between these two variables. Seasonal effects show a weak positive correlation of 0.19 with temperature, reflecting a mild influence of seasonal variations on temperature fluctuations. Strong negative correlations are observed between temperature and energy consumption, with values of -0.69 for hot water power and -0.60 for steam volume, indicating that warmer temperatures are associated with reduced energy usage for these metrics. A moderate negative correlation of -0.27 is noted between temperature and electricity energy consumption, suggesting a less pronounced but noticeable reduction in electricity use during warmer periods.

The correlation analysis presented in Table 3.3 serves as an exploratory tool to assess the linear relationships between environmental variables (e.g., temperature, humidity, solar radiation) and energy consumption metrics (electrical energy, hot water power, gas volume) on the UBC campus. While the correlation values were not directly embedded into the ANN architecture, they provided important guidance during feature selection, helping prioritize variables that exhibited stronger relationships with the target outputs.

It is important to note that all machine learning (ML) and deep learning (DL) models in this study were customized for specific prediction tasks, such as forecasting electrical energy or hot water power. As such, the choice of features, model hyperparameters, and validation strategies were tailored to the nature of each problem and data type. Therefore, discrepancies between the variables in Table 3.3 and those in the ANN model reflect deliberate optimization decisions based on model performance metrics (e.g., MAE, RMSE, R^2) rather than strict adherence to correlation rankings.

Additionally, while the numerical results including correlation coefficients and model weights are context-specific due to UBC’s unique energy infrastructure and climatic conditions, the methodological approach (i.e., correlation analysis for feature insight, model customization, and performance-based evaluation) is generalizable. Future researchers and practitioners can adapt the same framework to other urban or campus-scale energy systems by substituting site-specific data and retraining the models accordingly. This ensures that while the models are localized, the workflow and methodology remain transferable for broader energy analytics and forecasting applications.

Table 3.3 Correlation Analysis of Energy Consumption Metrics and Environmental Factors at UBC Campus

Metric Pair	Correlation	Interpretation
Electricity Energy and Hot Water Power	0.38 (Moderate Positive)	Increases in electricity usage are moderately associated with increases in hot water power consumption.
Electricity Energy and Gas Volume	0.25 (Weaker Positive)	Weaker positive relationship; changes in electricity usage have a less pronounced effect on gas volume.
Electricity Energy and Steam Volume	0.31 (Moderate Positive)	Moderate positive correlation; notable relationship between electricity usage and steam volume.
Hot Water Power and Steam Volume	0.75 (Strong Positive)	
Hot Water Power and Gas Volume	0.34 (Moderate Positive)	Moderate positive relationship between hot water power consumption and gas volume.
Gas Volume and Steam Volume	0.49 (Moderate Positive)	Moderate positive correlation; considerable relationship between gas and steam volumes.
Seasonal Effects and Temperature	0.19 (Weak Positive)	Weak positive correlation; mild influence of seasonal variations on temperature fluctuations.
Temperature and Hot Water Power	-0.69 (Strong Negative)	Strong negative correlation; warmer temperatures are associated with reduced hot water power usage.
Temperature and Steam Volume	-0.60 (Strong Negative)	Strong negative correlation; warmer temperatures are associated with reduced steam volume.
Temperature and Electricity Energy	-0.27 (Moderate Negative)	Moderate negative correlation; warmer temperatures are associated with a noticeable reduction in electricity use.

The moderate positive correlation ($r = 0.38$) between electricity consumption and hot water power usage likely reflects underlying operational and occupancy-driven factors. On the UBC campus, both systems are heavily influenced by building usage schedules, particularly during peak academic hours and colder seasons. For example, buildings with high occupancy, such as classrooms, laboratories, and residences, tend to have increased lighting, equipment use, and plug loads (affecting electricity) while simultaneously requiring space heating through hot water-based systems. Additionally, HVAC systems in some buildings utilize electric pumps and controls to circulate hot water, further linking the two energy streams. Therefore, this correlation suggests that shared operational patterns (e.g., occupancy, indoor comfort needs) drive concurrent increases in electricity and hot water energy demands.

The moderate positive correlation between hot water power consumption and gas volume arises from the direct dependency of the hot water heating system on natural gas. On the UBC campus, natural gas is the primary fuel used in central boilers to generate both steam and hot water for space heating and HVAC systems. When buildings require more thermal energy, particularly during colder periods, there is an increase in hot water circulation and temperature setpoints, which in turn increases the gas combustion rate to meet the heating load. This operational coupling explains why increases in hot water power demand are accompanied by higher gas volume consumption, resulting in the observed positive correlation in Table 3.3.

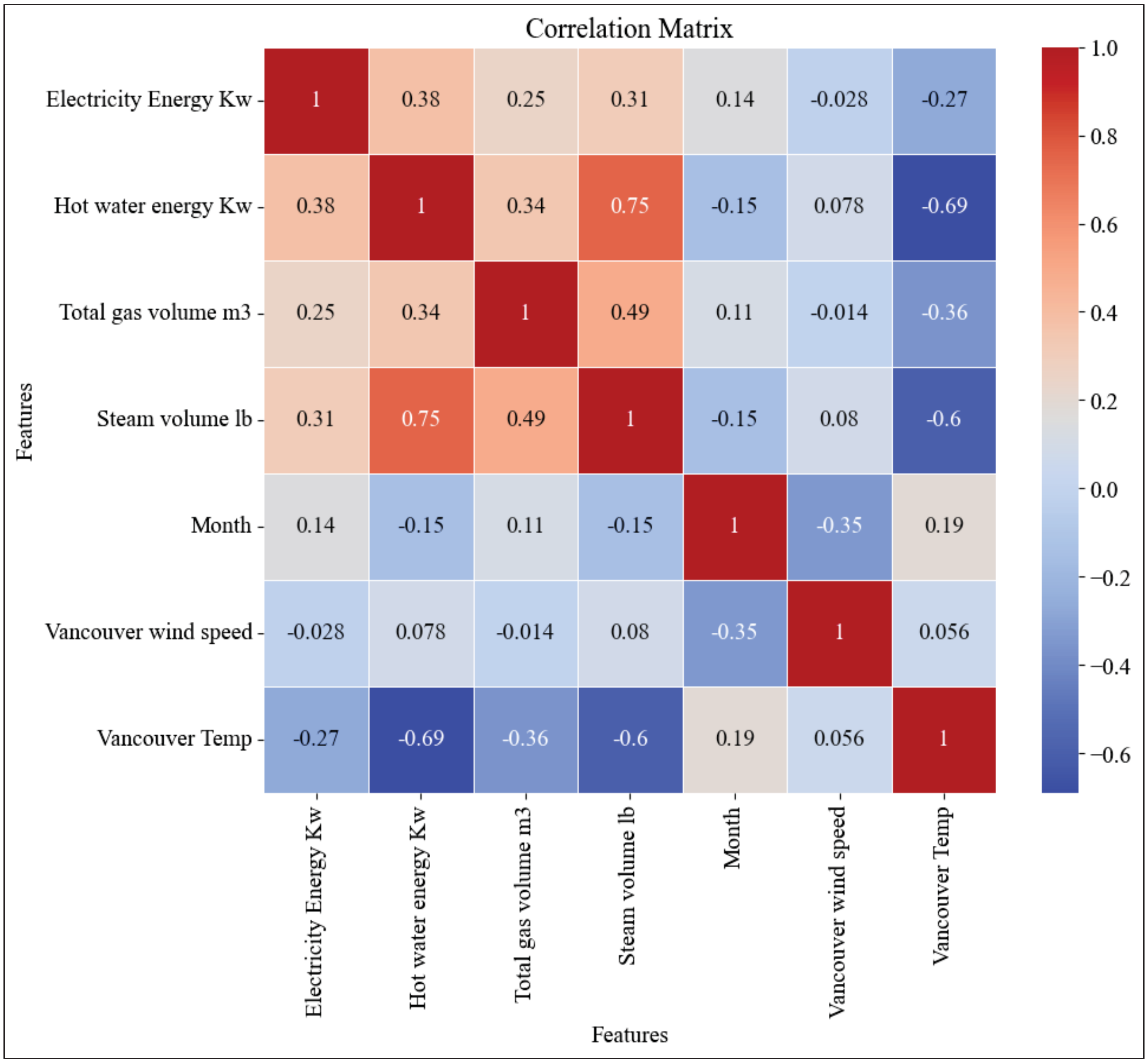

Figure 3.4 Correlation Matrix for weather and energy in UBC Campus

Assumptions were made in advance regarding the sign of the correlations. All relationships, such as those between hot water and steam volume or gas and electricity, were empirically derived from the dataset of the Campus in 2023. These relationships are inherently dynamic and campus-specific, particularly in response to changes in energy systems or equipment (e.g., fuel switching from gas to steam).

3.5.2 Electrical Energy

The analysis of various regression models for predicting electrical energy consumption based on a combination of meteorological variables and temporal features at the UBC Campus, which is shown in Figure 3.5, highlighted significant performance differences, emphasizing the critical role of model selection in energy usage. The Deep Neural Networks (DNN) model demonstrated the highest accuracy with a Mean Absolute Error (MAE) of 0.15 and a Coefficient of Determination (R^2) of 0.98, indicating its superior capability to capture complex patterns in the data. The Decision Tree Regressor also performed well, achieving an MAE of 0.18 and an R^2 of 0.92, reflecting good predictive accuracy. The Random Forest Regressor showed competitive performance with an MAE of 0.20 and an R^2 of 0.96, though slightly less accurate than the Decision Tree. In contrast, the Gradient Boosting Regressor had a higher MAE of 1.03 and a lower R^2 of 0.50, suggesting less effective prediction capabilities. AdaBoost performed the least effectively with an MAE of 1.22 and an R^2 of 0.34, indicating limited predictive accuracy. Linear Regression, Ridge Regression, Lasso Regression, Support Vector Regression, and K-Neighbors Regressor exhibited poorer performance overall, with MAE values above one and lower R^2 scores. This analysis underscores the advantage of employing advanced machine learning techniques, such as DNNs, for accurate energy forecasting, which is crucial for optimizing energy efficiency and resource management at the UBC Campus.

Table 3.4 Performance Comparison of Regression Models for Predicting Electrical Energy Consumption

Model	**Mean Absolute Error (MAE)**	**Coefficient of Determination (R^2)**
Deep Neural Networks (DNN)	0.15	0.98
Decision Tree Regressor	0.18	0.92
Random Forest Regressor	0.2	0.96
Gradient Boosting Regressor	1.03	0.5
AdaBoost Regressor	1.22	0.34

Linear Regression	>1	Lower values
Ridge Regression	>1	Lower values
Lasso Regression	>1	Lower values
Support Vector Regression	>1	Lower values
K-Neighbors Regressor	>1	Lower values

Table 3.4 provides a clear comparison of the performance metrics of various regression models, enabling direct assessment of their effectiveness in predicting electrical energy consumption.

The four correlations for electrical energy presented in Table 3.3 reflect pairwise linear relationships between individual input variables (e.g., temperature, wind speed) and the target output. In contrast, the MAE values reported in Table 3.4 correspond to multi-variable regression models, where all selected features were used as inputs. Therefore, the correlation values and the MAE metrics reflect different stages of analysis exploratory and predictive, respectively.

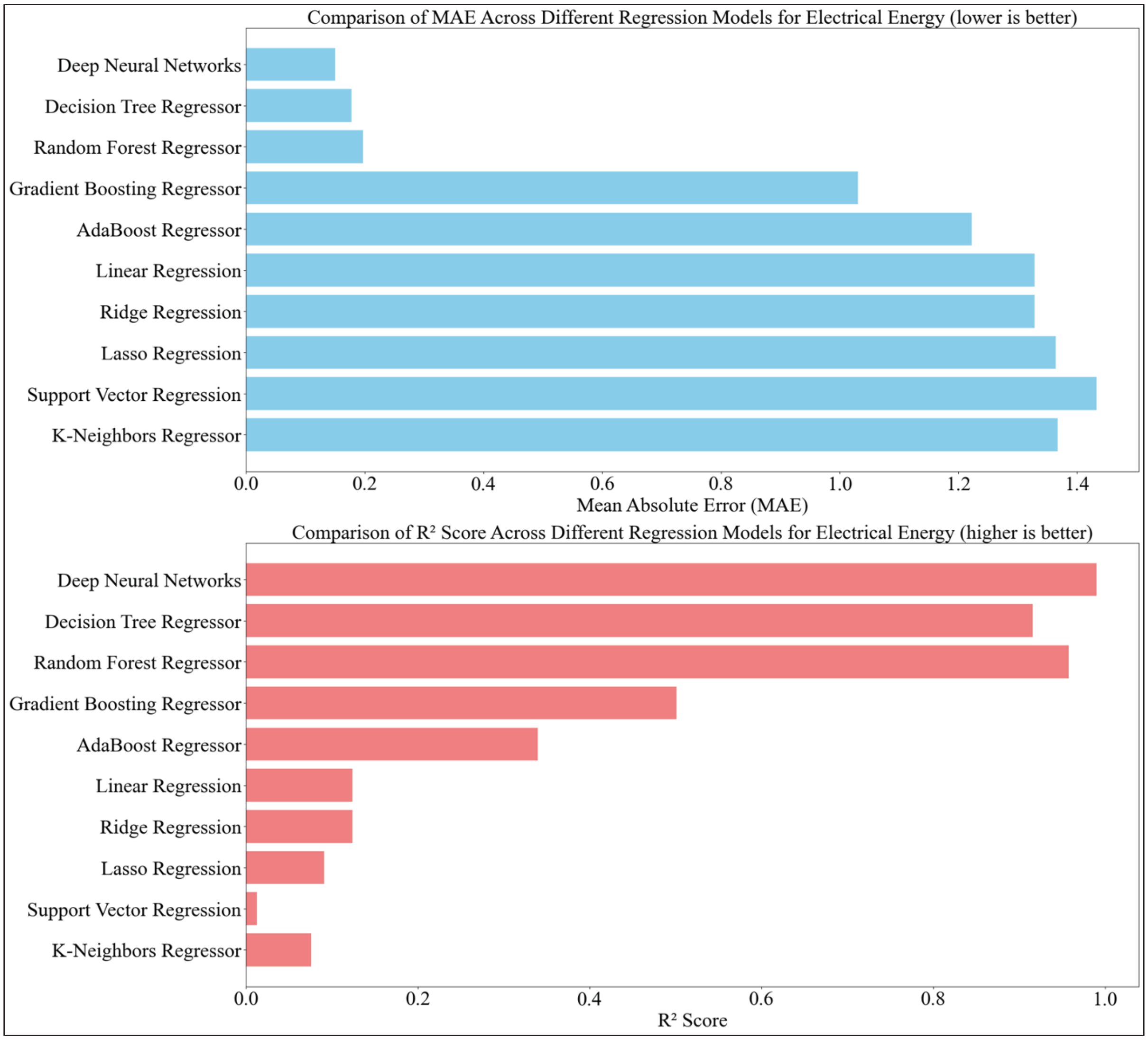

Figure 3.5 Comparison of MAE and R2 Score for Electricity Energy

3.5.3 Hot Water Power

The analysis of various models for predicting hot water power consumption on the UBC campus revealed significant differences in performance (Figure 3.6). The Neural Network model demonstrated strong predictive performance, with a Mean Absolute Error (MAE) of 3570.00 and a high R^2 value of 0.88, indicating its effectiveness in accurately estimating hot water power usage. The Decision Tree Regressor also performed well, with an MAE of 2038.71 and an R^2 of 0.89, reflecting its capability to capture patterns and trends in hot water power consumption. The Random Forest Regressor achieved the lowest MAE of 1852.85 and a high R^2 value of 0.91, showing its effectiveness in predicting hot water power consumption and managing heating systems efficiently.

In contrast, the Gradient Boosting Regressor had a higher MAE of 2660.06 and a moderate R^2, suggesting it may need further optimization for improved accuracy. AdaBoost Regressor exhibited the highest MAE of 3986.78 and a moderate R^2, indicating less effectiveness compared to the top-performing models. The linear models, including Linear Regression, Ridge Regression, and Lasso Regression, had an MAE of around 5298.00 and an R^2 of approximately 0.51, providing a baseline with moderate performance but not optimal for precise predictions. Support Vector Regression had the highest MAE of 8222.55 and the lowest R^2, indicating significant limitations in accurately modelling hot water power consumption. The KNeighbors Regressor showed reasonable capabilities with an MAE of 2413.37 and an R^2 of 0.88, performing comparably to the Neural Network model.

In summary, the Neural Network model emerged as the most effective for predicting hot water power consumption, followed closely by the Decision Tree and Random Forest models. These findings underscore the benefit of using advanced machine learning techniques for accurate hot water system management and optimization.'

Table 3.5 Performance Comparison of Models for Predicting Hot Water Power Consumption

Model	MAE	R^2
Neural Network	3570	0.88
Decision Tree Regressor	2038.71	0.89
Random Forest Regressor	1852.85	0.91
Gradient Boosting Regressor	2660.06	Moderate
AdaBoost Regressor	3986.78	Moderate
Linear Regression	~5298.00	~0.51
Ridge Regression	~5298.00	~0.51
Lasso Regression	~5298.00	~0.51
Support Vector Regression	8222.55	Lowest
KNeighbors Regressor	2413.37	0.88

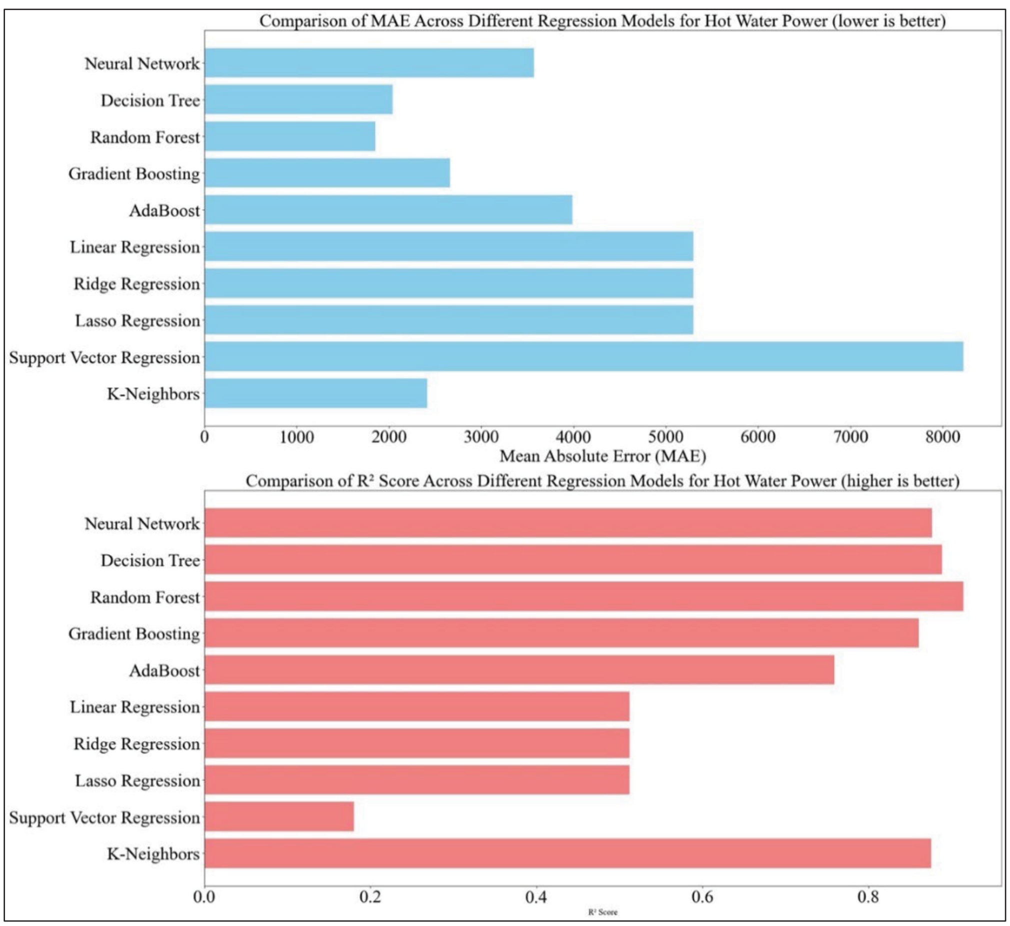

Figure 3.6 Comparison of MAE and R2 Score for Hot water power

3.5.4 Gas Volume

The analysis of various models for predicting gas volume consumption on the UBC campus revealed distinct differences in performance (Figure 3.7). The Neural Network model emerged as the top performer, achieving a Mean Absolute Error (MAE) of 28.82 and a high R^2 value of 0.96. This indicates its exceptional ability to capture the underlying patterns and variability in gas volume data, making it highly effective for optimizing gas usage. The Random Forest Regressor also showed strong performance, with an MAE of 33.51 and an R^2 of 0.91, making it a reliable tool for accurate gas volume forecasting. The Decision Tree Regressor demonstrated moderate predictive capabilities with an MAE of 37.17 and an R^2 of 0.86. While useful, it may need refinement to achieve the accuracy of the Neural Network and Random Forest models. In contrast, the Gradient Boosting Regressor exhibited a higher MAE of 57.84 and a moderate R^2 of 0.55, indicating some predictive power but also room for improvement. AdaBoost Regressor had the highest MAE of 67.17 and the lowest R^2 of 0.44, suggesting limited effectiveness and a need for further optimization. The linear models, including Linear Regression, Ridge Regression, and Lasso Regression, had a poor predictive performance with an MAE of around 82.00 and an R^2 of approximately 0.18, providing only a baseline for comparison. Support Vector Regression showed the highest MAE of 88.39 and the lowest R^2 of 0.13, reflecting significant limitations in accurately modelling gas volume consumption. The KNeighbors Regressor displayed reasonable performance with an MAE of 54.32 and an R^2 of 0.57, performing better than some models but not as effectively as Neural Networks or Random Forest.

The Neural Network model emerged as the most effective for predicting gas volume consumption, with the lowest MAE and highest R^2. The Random Forest Regression Model also performed well, providing accurate forecasts. Other models, such as Decision Trees, Gradient Boosting, AdaBoost, linear regression models, and Support Vector Regression, showed varying degrees of performance, with some needing further refinement to enhance accuracy. These findings underscore the effectiveness of advanced machine learning techniques for precise gas consumption forecasting and efficient fuel management.

Table 3.6 Performance Comparison of Models for Predicting Gas Volume Consumption

Model	MAE	R²
Neural Network	28.82	0.96
Random Forest Regressor	33.51	0.91
Decision Tree Regressor	37.17	0.86
Gradient Boosting Regressor	57.84	0.55
AdaBoost Regressor	67.17	0.44
Linear Regression	~82.00	~0.18
Ridge Regression	~82.00	~0.18
Lasso Regression	~82.00	~0.18
Support Vector Regression	88.39	0.13
KNeighbors Regressor	54.32	0.57

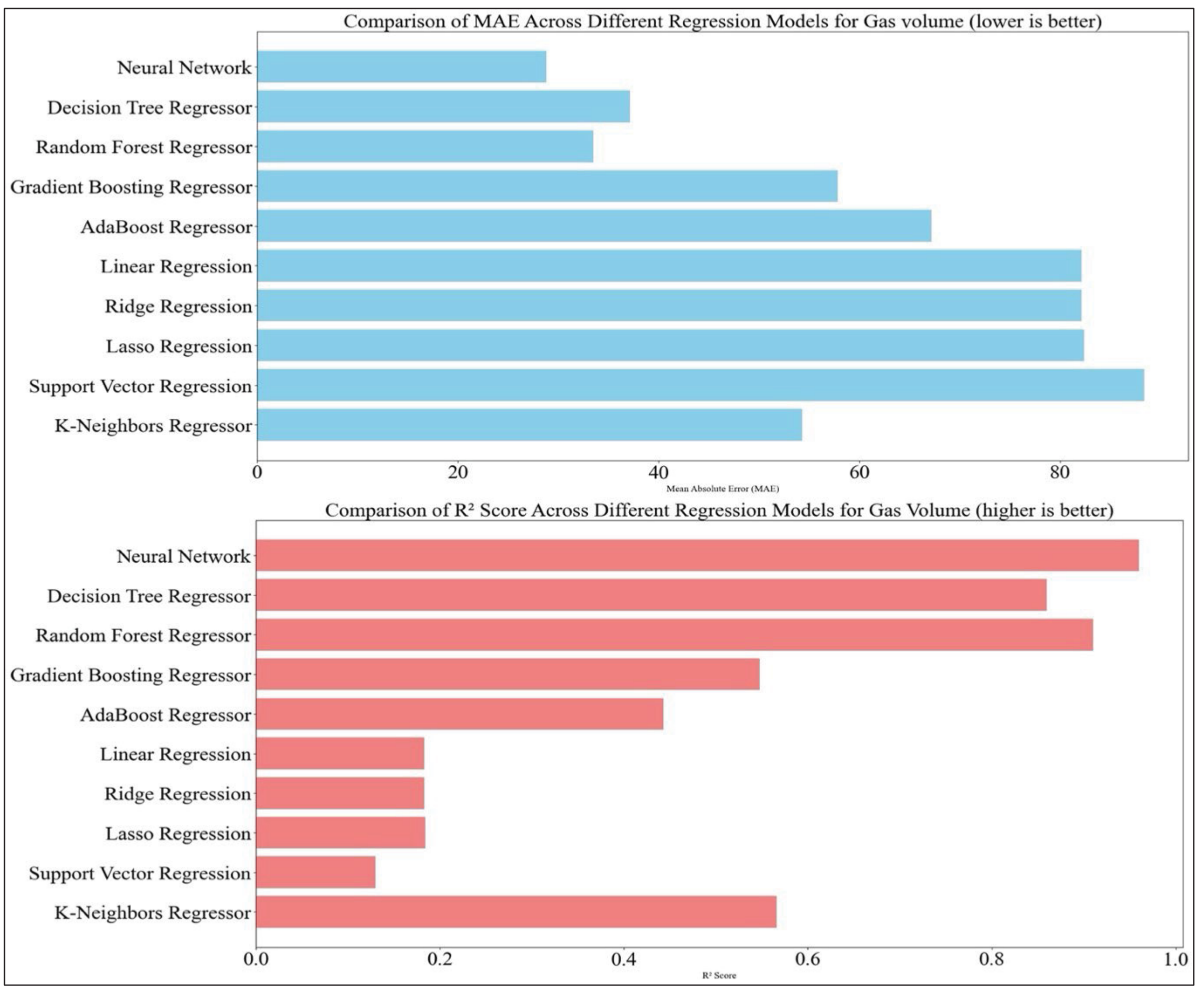

Figure 3.7 Comparison of MAE and R2 Score for Gas Volume

Synthesis of the predictive accuracy for the different data sets

The analysis of the R^2 values for various regression models applied to electrical energy, hot water power, and gas volume at the UBC Campus reveals significant differences in predictive accuracy (Figure 3.8):

1. Deep Neural Networks consistently demonstrate the highest R^2 values across all three energy forms, with an outstanding 0.98 for electrical energy, 0.88 for hot water power, and 0.96 for gas volume, indicating solid predictive performance.

2. Decision Tree Regressor and Random Forest Regressor perform well, particularly for hot water power (0.888 and 0.914, respectively) and gas volume (0.863 and 0.914, respectively). However, their performance for electrical energy is somewhat lower (0.916 and 0.958, respectively).

3. Gradient Boosting Regressor and AdaBoost Regressor show moderate performance, with R^2 values around 0.50 for electrical energy and gas volume. Still, they perform better for hot water power with R^2 values of 0.860 and 0.759, respectively.

4. Support Vector Regression (SVR) and K-Neighbors Regressor display the least predictive accuracy, with R^2 values significantly lower across all energy types, particularly for electrical energy, where the values are 0.013 and 0.076, respectively. These models show limitations in capturing the variability in the data.

5. Simpler models such as Linear Regression, Ridge Regression, and Lasso Regression exhibit notably lower R^2 values, particularly for electrical energy, with values approximately 0.124 for both Linear and Ridge Regression and even lower for Lasso Regression at 0.091. Their performance is similarly suboptimal for hot water power and gas volume, where the R^2 values hover around 0.512 for both Linear and Ridge Regression and are slightly lower for Lasso Regression.

Overall, the results indicate that complex models, particularly ensemble methods and deep learning, provide superior predictive accuracy for the energy consumption metrics at the UBC Campus. In contrast, more straightforward regression techniques and some specialized models like SVR and K-Neighbors struggle to effectively capture the data's variability.

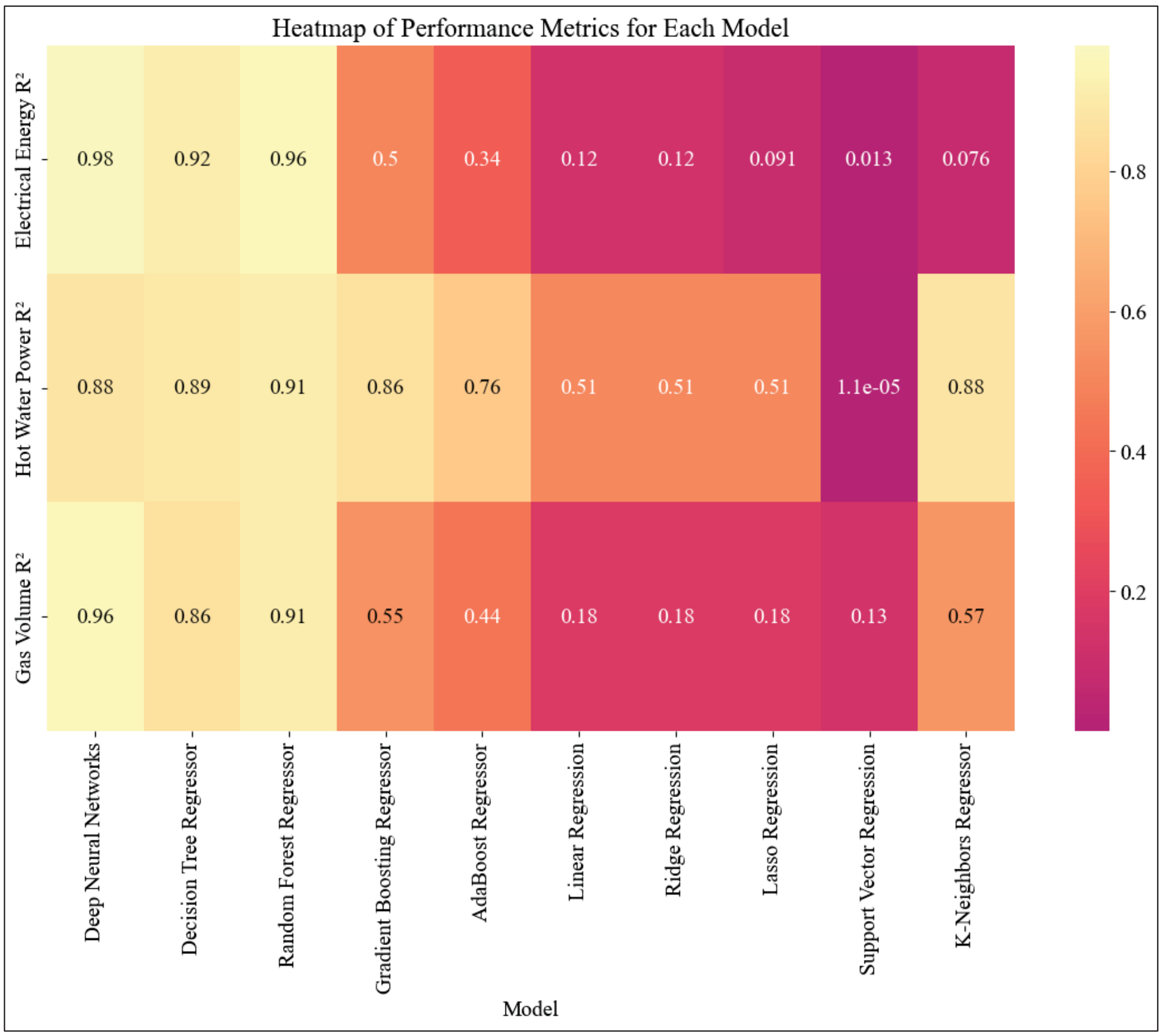

Figure 3.8 Heatmap of R2 squared

Figure 3.8 illustrates the training and test loss graphs showing the different prediction model trends for the different energy components data at the UBC Campus. The main observations are:

1. For the Electrical Energy Model:

Rapid Convergence: This model shows a quick decrease in both training and test loss during the early epochs, indicating effective learning.

Strong Generalization: The close alignment of training and test loss curves demonstrates that the model generalizes well to unseen data, avoiding overfitting and underfitting.

Balanced Fit: The lack of a significant gap between training and test loss suggests that the model accurately captures the complexity of electrical energy consumption without fitting noise in the data.

2. Hot Water Model:

Slower, Fluctuating Convergence: This model's loss curves show more fluctuation and a slower decrease, reflecting challenges in stabilizing it.

Initial Overfitting: Early in training, the model performs better on training data than test data, indicating overfitting.

Improved Generalization: Over time, the gap between training and test loss narrows, suggesting that the model improves generalization. However, the final model still shows some imbalance.

3. Gas Volume Model:

Gradual Convergence: This model exhibits relatively slow convergence, with a gradual decrease in loss throughout training, suggesting ongoing learning.

Overfitting: The persistent gap between training and test loss indicates that the model may be memorizing training data rather than learning general patterns, pointing to overfitting.

Potential for Improvement: To enhance generalization, the model could benefit from adjustments such as regularization, more data, or changes in the model architecture.

The strong performance and good generalization of the electrical energy model make it suitable for predicting electrical energy consumption, and we recommend continuing to use this model and monitoring its performance. For the hot water model, it is necessary to address the initial overfitting by applying techniques like early stopping or regularization and ensure that the model maintains balanced performance as it continues to learn. The gas volume model can be improved by exploring regularization methods, expanding the dataset, or adjusting the architecture to better capture patterns and reduce overfitting.

This analysis provides insights into each model's performance and offers guidance for refining approaches to achieve better predictive accuracy and more effective energy management at the UBC campus.

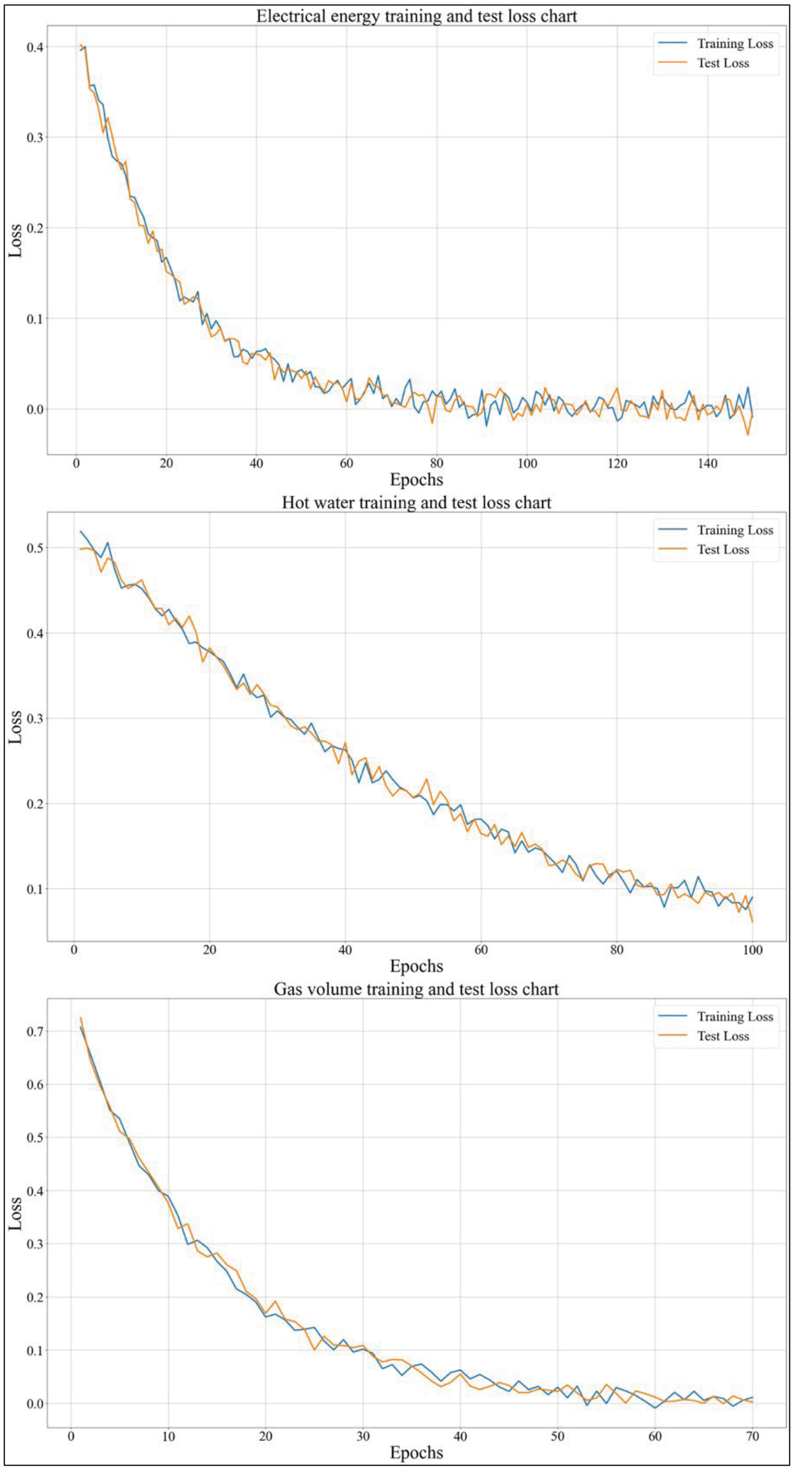

Figure 3.9 Training and Test Loss Trends for Energy Usage Prediction Models at UBC Vancouver Campus

3.6 Conclusion

In conclusion, this comprehensive analysis of energy consumption patterns at the UBC campus in Vancouver has revealed critical insights into the intricate relationships among various energy sources, environmental factors, and operational strategies. Through a robust correlation analysis, this study identified significant interdependencies between electricity consumption, hot water power, gas volume, and steam volume, emphasizing the need for an integrated, system-wide approach to energy management. Notably, the strong inverse correlations between Vancouver temperature and multiple energy consumption metrics suggest that temperature-responsive strategies could significantly enhance energy efficiency, particularly during warmer periods, aligning with existing research on adaptive energy management practices in temperate climates.

The evaluation of multiple regression models, particularly the superior performance of deep neural networks, highlighted their capacity to capture complex, nonlinear patterns in energy consumption data. These models consistently outperformed traditional methods, demonstrating high predictive accuracy and offering valuable, data-driven insights. The effectiveness of these machine learning techniques underscores their potential to optimize energy use, reduce wastage, and lower environmental impact, contributing to the growing body of literature that advocates for the application of advanced analytics in energy systems optimization.

These findings hold substantial implications for energy management at UBC and similar institutions. By understanding the complex interplay between energy sources, environmental variables, and operational conditions, stakeholders are empowered to implement more targeted, effective interventions that minimize energy consumption, promote sustainability, and generate cost savings. Furthermore, the success of deep neural networks in this study supports broader theoretical claims regarding the effectiveness of advanced machine learning algorithms in dynamic, multi-source energy systems, aligning with contemporary research in both energy management and artificial intelligence.

Looking ahead, future research should aim to further elucidate the causal relationships between energy consumption and external factors, evaluate the specific impacts of operational interventions, and explore the potential for integrating renewable energy sources into the campus's energy portfolio. By building upon the analytical framework and methodologies developed in this study, UBC can continue to lead in sustainable energy management while contributing to a growing body of work focused on reducing institutional environmental footprints and advancing theoretical understanding in the field of energy optimization.

Author Contribution

Conceptualization, A.S. and A.I.; methodology, A.S.; Coding, A.S.; validation, A.S. and A.I.; formal analysis, A.S.; investigation, A.S.; resources, A.S.; data curation, A.S.; writing original draft preparation, A.S.; writing review and editing, A.I.; visualization, A.S.; supervision, A.I.; project administration, A.S. and A.I.; funding acquisition, A.I.
All authors have read and agreed to the published version of the manuscript.

Funding

Data Availability Statement

The original contributions presented in the study are included in the article, further inquiries can be directed to the corresponding author.

Conflict of Interest

The authors declare no conflicts of interest.

CHAPTER 4

K-means and Agglomerative Clustering for Source-Load Mapping in Distributed District Heating Planning

Amir Shahcheraghian [a], Adrian Ilinca [a] , Nelson Sommerfeldt [b]

[a] Department of Mechanical Engineering, École de Technologie Supérieure, 1100 Notre-Dame West, Montreal, Quebec, Canada H3C 1K3

[b] Department of Energy Technology, KTH Royal Institute of Technology, SE-100 44 Stockholm, Sweden

Paper Published in *Energy Conversion and Management*:X, January 2025

Abstract

This study introduces a high-resolution, data-driven approach for optimizing district heating networks using source-load mapping, focusing on Stockholm as a case study. The methodology integrates detailed building energy performance data (2014–2022) with geographic data from the Swedish Survey Agency, employing advanced clustering techniques such as K-means Clustering, Agglomerative Clustering, DBSCAN, Spectral Clustering, and Gaussian Mixture Model (GMM) Clustering to identify optimal locations for distributed heat sources, including data centers, supermarkets, and water bodies. Quantitative results show that these environmentally friendly sources could supply 54% of Stockholm's total annual heat demand of 7.7 TWh/year, equating to 4.2 TWh from residual heat sources. Data centers contribute 0.48 TWh, water bodies provide 3.4 TWh, and supermarkets contribute 0.3 TWh annually. Economic analysis further reveals that 98% of residual heat sources are economically viable, with marginal costs of heat (MCOH) for data centers, supermarkets, and water bodies estimated at 12.7 EUR/MWh, 16.0 EUR/MWh, and 20.0 EUR/MWh, respectively well below the Open District Heating (ODH) market price of 22.0 EUR/MWh. The policy implications of these findings are profound. Policymakers can leverage this methodology to identify

economically viable heat sources, enabling the creation of regulations that incentivize the integration of distributed heat sources into existing district heating networks.

This can lead to reduced energy costs, enhanced sustainability, and more resilient energy systems. Practically, urban planners and energy utilities can use clustering insights to optimize the placement of new infrastructure, such as data centers, ensuring they are strategically located in high-demand zones. Furthermore, the study's methodology can be replicated in other urban contexts, offering cities worldwide a scalable tool for improving the efficiency and sustainability of their heating networks. These findings support the transition to low-carbon energy solutions and provide actionable recommendations for the long-term development of urban energy systems.

Keywords: Distributed District Heating, Clustering, Energy Performance Certificates, Economic Viability, Urban Energy Planning, Heat Source Allocation, Sustainability, Marginal Cost of Heat, Policy Implications.

4.1 Introduction

District heating networks offer significant opportunities to rapidly transform energy supply in buildings and cities. Changing the fuel sources for heating plants makes it possible to increase the proportion of renewable energy or decarbonize the heating supply with minimal intervention from building owners. Sweden is a global leader in this area, with approximately 500 cities or communities utilizing district heating networks that supply heat to 55% of the building area (Werner, 2017). Most heat is generated by combustion fueled by residuals from the forestry industry and municipal solid waste, with peaks often covered by fossil fuels (Agency, 2023).

Traditionally, heating networks are centred around large, centralized plants. However, Stockholm's Open District Heating market illustrates how distributed heat sources such as supermarkets, ice rinks, and data centers can recover and contribute heat that would otherwise be wasted (Exergi, 2024a). This system allows facilities with substantial cooling demands to sell their excess heat to the network, creating a unique prosumer market for heat similar to the prosumer model in the electricity sector. While heavy industry has long contributed to heating networks through specialized agreements, this marketplace opens participation to smaller, distributed stakeholders. High spatial and temporal resolution data are essential to effectively integrate distributed heat sources into district heating networks. The growing availability of high-resolution geographic information systems (GIS) data has led to numerous district heating planning studies. On the demand side, large cities such as London (London, 2024), Berlin (IFAF, 2024), and Helsinki (Helsinki, 2024) have mapped out their building energy demands. Studies in Switzerland (Unternährer et al., 2017) and China (Su et al., 2019) have mapped the potential of ground-source heat pumps.

The electrification of heat production demands hourly time resolution to account for fluctuations in supply profiles and temperature levels, both of which impact system efficiency. Additionally, electricity prices, which have become more volatile since the withdrawal of Russian natural gas from the European market, vary by the hour. High temporal resolution is also crucial for accurately modelling storage systems, especially in such a volatile pricing environment.

This study uses high-resolution source load mapping to develop a novel data-driven approach to district heating network planning during the prefeasibility stages. Unlike existing methods, this approach integrates detailed spatial data for both distributed heat demand and supply sources, identifying optimal locations for new heat sources and potential areas for heat recovery. This leads to more efficient and cost effective district heating infrastructure designs. Stockholm is used as a case study due to the availability of data, particularly the distributed heat-source dataset from Su et al. (Su et al., 2021), on which this research is based. The study's objectives include analyzing how existing distributed resources can be optimally allocated to meet the city's heat demand, identifying the best locations for new data centers that can serve as significant heat sources, assessing the economic feasibility of utilizing heat from these distributed sources, and developing a comprehensive spatial data-base that includes detailed building energy performance metrics across multiple years (2014-2022) along with extensive geographical data. By addressing these issues, the study aims to provide actionable insights that enhance the sustainability and cost-effectiveness of urban heating solutions.

This study addresses several significant research gaps in the district heating field, particularly in integrating distributed heat sources into urban energy systems. First, traditional district heating studies have largely focused on centralized heat production, overlooking the potential of distributed sources such as data centers, supermarkets, and water bodies.

As a result, the significant untapped potential of residual heat from these sources has often been underutilized. Another critical gap is the reliance on coarse resolution spatial and temporal data in previous studies, which limits the ability to capture urban heat supply and demand variability and dynamics. The lack of high-resolution data hampers the accuracy of supply-demand matching and the identification of optimal resource allocation.

Additionally, while the technical potential of distributed heat sources has been well documented, there has been insufficient research exploring their economic viability in the context of current market conditions. This gap leaves policymakers without the necessary insights to make informed decisions about integrating these sources into existing infrastructure. Furthermore, previous studies have not fully leveraged advanced analytical approaches, such as machine learning algorithms. In particular, unsupervised learning methods, including K-means, DBSCAN, and Gaussian Mixture Models, have been underutilized in analyzing the complex spatial and temporal patterns of heat supply and demand. The lack of these advanced analytical techniques has constrained the ability to handle noise, non-linear distributions, and the scalability needed in urban energy analysis. Another overlooked aspect is the role of data centers, which, despite their potential as significant heat sources, have often been underrepresented in district heating studies due to outdated or insufficient data. The limited use of localized analysis tailored to specific urban contexts has further hindered the recognition of their potential. Moreover, many studies fail to provide methodologies that can be replicated across different urban contexts, limiting the broader applicability of their findings. This lack of replicability impedes the widespread adoption of sustainable district heating solutions globally. Finally, a critical gap exists between academic insights and the practical needs of urban planners and policymakers. Previous research has often failed to offer actionable recommendations, hindering the translation of academic findings into real-world applications.

By addressing these gaps, this study provides a high-resolution, economically grounded, and methodologically robust framework for integrating distributed heat sources into district heating networks, offering valuable insights for both researchers and urban planners.

4.2 Literature Review and Background

Blanco et al. (Blanco Bohorquez et al., 2023) introduced a novel data-driven methodology for urban energy analysis by classifying urban areas into sixteen distinct morphological units called Urban Energy Units (UEUs). This approach utilizes open-source data and machine learning techniques, including a random forest model to address missing building data and a decision tree model for UEU classification. This data-driven method facilitates the creation of modular energy districts within cities, leading to more targeted and effective energy planning. Applied in Oldenburg, Germany, the methodology demonstrates the practical benefits of their classification system and underscores the importance of data-driven approaches in urban energy analysis for informed decision-making. Their study not only tackles the issue of missing data but also lays the groundwork for future research to validate and expand the classification framework to other urban areas. This work aligns with our studies, particularly in its application to district heating systems, highlighting the importance of data-driven approaches for informed energy planning.

Zhang et al. (Zhang et al., 2018) tackle the critical issue of rising energy consumption in data centers, which has surged dramatically in recent years. The paper comprehensively reviews the latest thermal management techniques and evaluation metrics for data centers, discussing energy conservation strategies and safe operation practices. It covers critical advancements in cooling technologies, innovative energy-saving methods such as free cooling and heat recovery, and optimization techniques to enhance system efficiency. The paper emphasizes the significance of thermal evaluation metrics in maintaining equipment safety and energy efficiency. It concludes by calling for further research and strategic development in the thermal management of data centers.

Wahlroos et al. (Wahlroos et al., 2018) explore the potential of reusing waste heat from data centers in Nordic countries, highlighting both the challenges and opportunities. The paper addresses issues such as the lack of transparency in business models between district heating and data center operators and waste heat reuse's economic and environmental implications. Their proposed 8-step change process provides a structured approach to overcoming barriers, aligning closely with our project’s focus on waste heat management in data centers and district heating systems. The review emphasizes the importance of understanding energy efficiency metrics and economic considerations directly relevant to our project's objectives.

Davies et al. (Davies et al., 2016) investigated the potential for reusing waste heat from data centers in London, focusing on using heat pumps to upgrade waste heat for district heating networks. The study reveals significant energy, carbon, and cost savings potential, especially when data center waste heat is integrated into district heating systems with eligibility for Renewable Heat Incentive (RHI) payments. It aligns potential waste heat sources with the heat demands of various London districts, demonstrating substantial energy and carbon savings benefits.

Cichowicz et al. (Cichowicz & Jerominko, 2023) compared Energy Performance Certificate (EPC) data calculation methods with actual energy consumption to evaluate energy efficiency in multi-family residential buildings. Their findings revealed a 14.5% difference in Practical Energy Consumption (PEC) indices and a 14.7% difference in Final Energy (FE) indices between the two methods. While both methods showed comparable performance, the reliance on user behaviour in the consumption method raised concerns about its reliability across different building types. This study highlights the need for standardized procedures considering building specific characteristics in EPC-based energy assessments.

Zhang et al. (Y. Zhang et al., 2022) explore the evolution of district heating and cooling (DHC) systems in Sweden, focusing on adapting these systems to future changes in demand profiles and renewable energy supplies. Using a generalized methodology framework that integrates future changes, various operational scenarios, and system design optimizations, the study concludes that a fifth-generation district heating and cooling (5GDHC) system is the most economically viable option in a future scenario with low-energy building stock and increased cooling demand (Stef Boesten, 2019). The study also highlights the significant impact of electricity prices on the cost-efficiency of 5GDHC and ultra-low temperature district heating and cooling (ULTDHC) systems, particularly with a 50% share of wind power in the national grid. The methodology presented can be applied to similar systems to enhance understanding of system transitions.

Huang et al. (Huang et al., 2020) investigate the dual role of data centers in district energy systems, where they act as both consumers and producers of energy. The study reviews the integration of renewable energy and the recovery and reuse of waste heat in data centers, emphasizing the importance of optimizing both upstream and downstream processes. It discusses how upstream integration involves procuring and managing renewable energy sources, such as solar and wind, for data center operations. At the same time, downstream waste heat utilization refers to managing and repurposing waste heat for district heating systems. The findings indicate that while integrated global controls for managing energy production, operation, and waste heat generation are still developing, regional climate studies are crucial for optimizing these integrations. The study also highlights the development of global energy metrics as essential for quantifying data center performance and providing a comprehensive approach to sustainable energy use within data centers.

Su et al. (Su et al., 2021) focus on decarbonizing Stockholm's district heating sector to achieve net-zero emissions by 2040. They use an integrative GIS-based analysis to map clean, non fossil fuel heat sources within Stockholm, achieving high-resolution mapping and addressing data availability challenges. Their results show that the city has abundant clean heat sources, including water bodies and data centers, capable of covering 100% of the district's heating net annual energy needs. The study identifies clusters of heat sources for prioritized exploitation. It provides a method pipeline applicable to other cities transitioning to clean district heating, emphasizing the importance of strategic planning based on local heat source availability.

Ogliari et al. (Ogliari et al., 2022) proposed a methodology integrating clustering and neural networks to enhance day-ahead thermal load forecasting for a District Heating (DH) system in Northern Italy.

The study comprehensively addressed data preprocessing, variable correlation analysis, clustering, and forecasting. Three clustering techniques, K-means, Hierarchical Agglomerative Clustering (HAC), and DBSCAN were evaluated using the Caliński-Harabasz and Silhouette indexes, identifying three optimal clusters while excluding outliers, but no single superior method emerged. Forecasting was explored using three strategies: training a single neural network for all utilities, per cluster (based on HAC), and substation. Results indicated that training a neural network for each substation yielded the highest accuracy. However, for larger systems, a cluster-based approach was recommended to balance accuracy and computational efficiency. The authors emphasized the need for further research with expanded datasets and additional substations to validate and generalize their findings.

Lumbreras et al. (Lumbreras et al., 2021) highlights the critical role of the building sector, consuming approximately 40% of primary energy in the European Union, and the increasing focus on energy efficiency driven by EU directives. District Heating (DH) networks, covering 13% of EU building heat loads, are evolving towards 4th Generation District Heating (4GDH), characterized by low-temperature heat distribution and renewable energy integration. The study identifies a gap in the literature regarding the application of unsupervised clustering techniques to heating demand data, unlike the extensive application in electricity demand analysis. The authors propose a multistep clustering framework using density-based clustering

for outlier detection and k-means for identifying heating consumption patterns. The methodology reveals the interpretative challenges of clustering, as optimal classifications vary based on cluster validation indices (e.g., K = 3). While the framework offers valuable insights into demand patterns, its replicability remains limited. Future work aims to explore correlations between demand patterns and climatic or calendar variables and extend the framework's application to additional buildings within the DH network to enhance gen
eralizability and applicability.

4.3 Methodology

The proposed method involves aggregating spatially labelled heat energy demand data with distributed heat sources, clustering these demands to map supply, and identifying optimal locations for new heat sources. Additionally, it includes calculating levelized heating costs by cluster. Energy demand data is sourced from the Swedish Housing Authority's (Boverket) Energy Performance Certificates (EPC) database and geographically located using data from the Swedish Survey Agency (Lantmäteriet) (Boverket – the Swedish National Board of Housing, Building and Planning., 2024; The Swedish Survey Agency, 2024). This database is merged with the locations of existing data centers, supermarkets, and water bodies, providing a comprehensive, multi-dimensional perspective for practical spatial analysis (Su et al., 2021).

The integration process synchronizes building specific energy performance metrics with their geographical coordinates, ensuring that each data point accurately reflects both location and energy profile. Advanced clustering techniques, including k-means and agglomerative methods, are then employed to segment the region based on heat demand characteristics (al, 2024a, 2024b). These clusters are analyzed to identify central nodes representing optimal locations for new data centers, considering their potential contributions to district heating systems.

The methodology ensures the optimal placement of new data centers within these identified clusters using quantitative metrics derived from k-means and agglomerative clustering analyses which mean that this study are looking for new potential locations for the existing network. These analyses leverage heat demand data from all buildings in Stockholm City. This approach allows for evaluating data center locations' proximity to high-heat demand areas and their potential for integration into heat recovery systems. It aligns with existing urban elements, such as supermarkets and water bodies, to enhance sustainable energy management. Ultimately, this method provides a data-driven framework for strategic urban planning and resource management, facilitating the development of district heating solutions that are both economically viable and environmentally sustainable.

The methodology for this analysis involves using Python for data processing and machine learning, with key libraries such as Pandas, NumPy, Scikit-learn, and Matplotlib. The simulation applies various clustering models, including K-means Clustering, Agglomerative Clustering, DBSCAN, Gaussian Mixture Models (GMM), and Spectral Clustering, to group the data based on features like heat demand, energy usage, and geographical locations. Visualizations are enhanced by overlaying geographical scatter plots with a background image, providing context for the clustering results. The dataset for Stockholm spans multiple years, with EPC (Energy Performance Certificate) data sizes for each year as follows: 10,368 data points for 2014, 39,320 for 2018, 42,566 for 2019, 28,830 for 2020, 14,304 for 2021, and 7,896 for 2022. The total dataset size across these years amounts to 143,284 data points. This dataset includes detailed information on energy performance, property designation, and geographical coordinates, which are essential for clustering and analysis in the simulation. K-means Clustering serves as the primary solver in this environment, along with other unsupervised learning algorithms such as Agglomerative Clustering, DBSCAN, Gaussian Mixture Model (GMM), and Spectral Clustering. These clustering methods are employed to categorize the data based on key attributes, including heat demand, energy usage, and geographical locations.

The dataset is comprised of both categorical and numerical attributes. To handle missing values, imputation and interpolation methods are applied. Additionally, the data is normalized to ensure it is properly scaled for effective clustering analysis.

4.3.1 Processing of EPC data with Swedish survey agency data

The EPC data was integrated with geographic information from the Swedish Survey Agency (Lantmäteriet) to identify optimal locations for future data centers based on heat demand characteristics (Boverket – the Swedish National Board of Housing, Building and Planning., 2024; The Swedish Survey Agency, 2024).

The data collection encompassed a wide array of energy metrics, categorized by energy types used for heating and domestic hot water (e.g., district heating, heating oil, natural gas, firewood, and various forms of electricity, including several types of heat pumps). These metrics were complemented by primary and normalized energy use figures for buildings, yielding a year-corrected value known as the Energy Index. Additional details included each building's address, complexity, category, type, construction year, and structural specifics, such as the number of heated basement floors, above-ground floors, stairwells, and residential apartments.

During integration, the energy performance data for each building was geocoded using primary and postal addresses, ensuring precise mapping. This accuracy is critical for practical spatial analysis, which underpins energy demand mapping and site selection for data centers.

In the data processing phase, efforts were made to standardize and normalize the data across different energy types and measurement units, facilitating uniform analysis. Rigorous data cleansing was conducted to correct inconsistencies, impute or exclude missing values based on their significance, and eliminate duplicate records. This involved removing whitespace from string fields for consistency, extracting non-numeric prefixes from property designations and address identifiers to standardize these fields, and handling missing values by removing rows with null entries. Additionally, coordinates were added based on property designations, and categorical variables were encoded using one-hot encoding. These steps ensured the dataset's integrity and reliability for subsequent analysis.

Through these processes, a robust and detailed spatial database was constructed to identify high-heat demand areas and determine the most strategic locations for future data center installations. This database not only supports detailed spatial analysis but also forms the foundation for the advanced clustering techniques used in the later stages of the study.

4.3.2 Heat demand Clustering using k-means and agglomerative clustering.

In this study phase, the geographical area was segmented into clusters based on buildings' heat demand characteristics using k-means and agglomerative clustering techniques.

The Elbow Method was utilized to determine the optimal number of clusters for k-means clustering. This method plots the sum of squared distances (inertia) from each point to its assigned cluster center against the number of clusters. As the number of clusters increases, inertia decreases. However, there is a point where the rate of decrease sharply slows, forming an "elbow" shape on the plot. This "elbow" indicates the optimal number of clusters, balancing the trade-off between underfitting and overfitting. By employing the Elbow Method, the segmentation was ensured to be both meaningful and efficient for further analysis (Rushirajsinh, 2023).

Subsequently, four centroids were selected within the dataset for the k-means clustering process. Each data point, representing the heat data of a building, was assigned to the nearest centroid based on the Euclidean distance between the data point's features and the centroid's features (www.geeksforgeeks.org). This assignment and recalculation of centroids were iteratively performed until the centroids stabilized, marking the algorithm's convergence. As a result, the dataset was divided into four distinct clusters, each representing an area with homogeneous heat demand characteristics.

Concurrently, agglomerative clustering was employed as a hierarchical method, starting with each data point as its cluster. Clusters were merged based on their similarity and assessed using

Ward’s method, which aims to minimize the variance within a cluster. This merging process continued until all data points were consolidated into four major clusters (Wijaya, 2019).

4.3.3 Allocation of heat sources to clusters

In the next step, k-means and agglomerative clustering were employed again, given their effectiveness in grouping data based on similarities, to identify clusters of Stockholm’s heat demands. To allocate existing renewable energy sources, Stockholm’s heat demands were segmented into ten districts using the k-means clustering algorithm, accessed via the scikitlearn library (www.geeksforgeeks.org).
The methodology begins by verifying and preparing the dataset, which includes critical analysis parameters. Data anomalies, such as infinite values or missing data points within the heat demand metric, are replaced with a statistically representative figure typically the mean or median to maintain data integrity. The heat demand data undergo a normalization process, ensuring a uniform scale for comparative visualization. The size of the visual markers is proportionally scaled to reflect the heat demand, enabling an accurate spatial representation of heat intensities.

For the clustering analysis, geographic coordinates are extracted to delineate the urban landscape into ten distinct zones. This segmentation is achieved through an iterative clustering technique that organizes the space based on proximity and similarity in heat demand, ensuring that the identified clusters are stable and replicable in subsequent analyses.

Furthermore, the algorithm calculates the central points of these clusters, pinpointing critical areas of heat concentration. The final visualization includes a legend correlating to the clusters, with the geographic scope of the map meticulously adjusted to encompass the Stockholm region of interest. This comprehensive display of the city's heat demand landscape is an indispensable tool for urban planning and the efficient allocation of energy resources.

4.3.4 Marginal cost of heat

The economic impact of waste heat recovery is evaluated using the marginal cost of heat (MCOH), primarily driven by electricity prices and limited to operational costs. This focus on operational costs is due to the relatively minor contribution of capital costs to the overall life cycle cost. The capital costs for heat recovery equipment are conservatively estimated at €1.5 million per MW of cooling capacity (Antal et al., 2019; Murphy & Fung, 2019; Oró et al., 2019), representing about 3% of the operational cost for a single year. Over a 20-year lifespan with a 10% discount rate, capital costs account for only 0.3% of the total life cycle cost, which falls well within the range of economic uncertainty.

Although capital costs are not insignificant, they are less critical in this context. Previous studies have indicated that economic returns for data center (Oró et al., 2019) and supermarket (Giunta & Sawalha, 2021) owners are uncertain, partly due to the variable value of heat throughout the year and the network owner's willingness to pay for it (www.geeksforgeeks.org). This study adopts the perspective of the district and city, viewing waste and environmental heat as part of a resource portfolio where the lowest marginal costs determine the merit order. For cooling devices and heat pumps, the MCOH is entirely influenced by the coefficient of performance (COP) and the price of electricity (pel), as represented by Equation (4.1). Maintenance costs, like capital costs, are also considered negligible and are therefore excluded from the analysis.

$$MCOH = \frac{p_{el}}{COP} \; [€/MWh] \tag{4.1}$$

Electricity prices and COPs fluctuate throughout the year, as does the value of heat, making detailed hourly simulations the most effective method for assessing cost effectiveness at any given time (Giunta & Sawalha, 2021; Sintong, 2023). However, high resolution time simulations are beyond the scope of this study. Instead, an economic performance map is created to capture a range of COPs and electricity prices.

This map includes cost ranges for data centers, supermarkets, and water bodies, comparing them to Stockholm's district heating prices at retail and wholesale levels. Wholesale prices are derived from Stockholm Exergi's 2024 base retail prices (Exergi, 2024b) and their Open District Heating (ODH) market (Raka Adrianto et al., 2018).

Equation (4.2) provides the formula for calculating the weighted MCOH for each district or cluster. Each cluster is denoted as i, and the MCOH of each heat source (with each technology denoted as j) is multiplied by the heat supply for each source in the cluster portfolio (Qj,i), then divided by the total heat demand of the cluster (Qi). This analysis reveals spatial differences in heating costs across the city, which can significantly influence urban development, particularly in cities lacking district heating networks.

$$MCOH_i = \frac{\sum \left({}^{p_{el,j}}/_{SCOP_j} \right) Q_{j,i}}{Q_i} \ [€/MWh] \tag{4.2}$$

4.3.5 Hyperparameter Selection and Justification

The hyperparameters listed in Table 4.1 are critical for selecting the optimal values for various clustering algorithms used in the study. In the case of K-Means (n_clusters = 4), this value was chosen based on the assumption that the data can be divided into distinct groups, such as high-demand versus low-demand heat zones. Selecting four clusters enables the identification of several types of demand and supply locations, representing different heating demand profiles across the city. The Elbow Method was employed to determine the ideal number of clusters by evaluating the sum of squared distances within clusters. A smaller number of clusters, such as two or three, would oversimplify the data, while more clusters could overfit the model and introduce noise, making four clusters an ideal balance for spatial clustering in district heating. Similarly, Agglomerative Clustering (n_clusters = 4) was selected to partition the dataset into distinct regions of heat supply and demand, with the number of clusters adjusted based on the dendrogram, which visually represents hierarchical relationships. Choosing fewer clusters could obscure subtle spatial patterns, while more clusters might increase complexity and risk overfitting, making four clusters a reasonable choice.

Table 4.1 Hyperparameter Selection for Clustering Algorithms

Algorithm	Hyperparameter	Value/Range	Description
KMeans	n_clusters	4	Number of clusters to form.
Agglomerative Clustering	n_clusters	4	Number of clusters to form.
DBSCAN Clustering	eps	[0.1, 0.3, 0.5, 0.7, 1.0]	The maximum distance between two samples for them to be considered as in the same neighborhood.
	min_samples	[3, 5, 10, 15]	The number of samples in a neighborhood for a point to be considered as a core point.
Spectral Clustering	n_clusters	4	Number of clusters to form.
	random_state	42	Seed used by the random number generator for reproducibility.
	affinity	'nearest_neighbors'	Metric used to compute the similarity matrix.
Gaussian Mixture Clustering	n_components	4	The number of mixture components (clusters).
	covariance_type	'full'	Type of covariance matrix. Options: 'full', 'tied', 'diag', 'spherical'.
	reg_covar	1.00E-04	Regularization to ensure covariance matrices are positive semi-definite.
	max_iter	500	Maximum number of iterations.
Data Preprocessing	imputer_strategy	'mean'	Strategy to use for imputing missing values.
	scaler	StandardScaler	Standardizes the data by removing the mean and scaling to unit variance.

For DBSCAN (eps = [0.1, 0.3, 0.5, 0.7, 1.0], min_samples = [3, 5, 10, 15]), the parameter eps controls the maximum distance between two points to be considered part of the same neighbourhood. A range of values from 0.1 to 1.0 was selected to explore different neighbourhood sizes, which is crucial for identifying tightly packed clusters (small eps values) or more spread out ones (larger eps values). The min_samples parameter determines the minimum number of points required to form a "core" point in a neighbourhood. Smaller values (e.g., 3) lead to the identification of more clusters, while larger values (e.g., 10 or 15) focus on denser, more meaningful clusters. This approach ensures that the algorithm can handle spatial data with varying densities, such as urban heat sources, and avoid noise. SpectralClustering (n_clusters = 4, affinity = 'nearest_neighbors', random_state = 42) was chosen to focus on local spatial structures by using the 'nearest_neighbors' affinity, which is critical for clustering heat demand or supply based on geographical proximity.

Fixing the random state ensures reproducibility of the results, which is essential for research purposes. The number of clusters affects the granularity of the identified patterns, influencing the resolution of heat demand-supply matching. For GaussianMixture (n_components = 4, covariance_type = 'full', reg_covar = 1e-4, max_iter = 500), the number of components was set to 4 to model the heat sources in Stockholm as a combination of four distinct distributions, each representing different heating profiles. The 'full' covariance type was selected to allow each component to have its own covariance matrix, offering maximum flexibility in modelling the data. Regularization (reg_covar = 1e-4) ensures the covariance matrices are positive semidefinite, which enhances model stability, especially with sparse data. The max_iter parameter was set to 500 to allow the algorithm sufficient time to converge to an optimal model. Fewer components may simplify the model, losing important detail, while more components could overfit the data.

Data preprocessing choices included imputing missing values using the mean strategy (imputer_strategy = 'mean') and standardizing the data with StandardScaler. The mean imputation was chosen because it is a simple and effective strategy for handling missing data, particularly when the data is missing at random. Standardizing the data ensures all features contribute equally to the clustering process by eliminating bias from features with larger ranges, such as geographical coordinates or temperature values. Other imputation strategies or scaling methods could be considered, but standard scaling is typically ideal for clustering tasks, as it normalizes the influence of each feature.

Overall, the choice of hyperparameters is crucial for accurately mapping and clustering heat supply and demand sources in Stockholm. Variations in cluster numbers, neighbourhood sizes, and covariance models impact the flexibility, sensitivity, and scalability of the clustering techniques. By selecting appropriate values, the study ensures that the results reflect meaningful patterns while avoiding overfitting or underfitting.

The use of multiple clustering techniques KMeans, AgglomerativeClustering, DBSCAN, and GaussianMixture enables the adaptation to different data characteristics, such as noise, density, and distribution shapes, providing a robust framework for integrating distributed heat sources into district heating networks.'

4.3.6 Methods Limitation

While the proposed methodology offers a comprehensive framework for optimizing heat resource allocation in urban environments, it has certain limitations. Any inaccuracies, missing entries, or outdated information can directly impact the clustering results and the overall accuracy of heat demand mapping. Additionally, the methodology assumes standardization across various energy types and units, which may not fully capture regional variations or temporal fluctuations in heat demand and supply. The clustering techniques used, such as K-means and agglomerative clustering, depend on the selection of hyperparameters, like the number of clusters, which can be somewhat subjective and may not fully reflect the complex spatial dynamics of urban heat demand.

Another significant limitation is the simplified economic model used for evaluating the marginal cost of heat (MCOH). By focusing primarily on operational costs and assuming that capital and maintenance costs are negligible, the methodology may overlook important economic factors influencing the feasibility of integrating new heat sources. Moreover, the method does not account for practical challenges such as policy constraints, stakeholder engagement, and the adaptability to real-time changes in heat demand or supply. These factors can significantly impact the implementation of the proposed solutions in real-world urban settings. Addressing these limitations is essential for enhancing the robustness and applicability of the methodology in diverse urban contexts.

4.4 Results

In this research, the marginal cost of environmental heat was assessed locally to determine its impact on the merit order of heat deployment and the resulting cost to the city. The heat demand across Stockholm was mapped using data from the Energy Performance Certificate (EPC) database, which was integrated with property locations provided by the Swedish Land Survey (Lantmäteriet).

After removing duplicate property description entries (known as 'fastighetsbeteckning' in Swedish), the number of unique properties in the EPC and Lantmäteriet datasets are 31,389 and 61,952, respectively. When the EPCs are filtered for those connected to district heating (approximately 52% of all buildings and 92% of the floor area in Stockholm) and cross-referenced for matching property descriptions, there are 14,103 entries. This is 86% of all DH-connected properties. However, the total demand is found to be 7.7 TWh/year and agrees with prior studies (Levihn, 2017). The breakdown by year of record for EPCs and matching properties with coordinates are given in Table 4.2. and shown spatially in Figure 4.1.

Table 4.2 Unique total property counts for EPCs and those with coordinates, by year

Year	Total EPCs	EPC and Coordinates	Match Rate
2014-2017	5583	1999	36%
2018	4258	3324	78%
2019	4931	2651	54%
2020	4867	2715	56%
2021	5696	1886	33%
2022	6054	1528	25%
Totals	31,389	14,103	45%

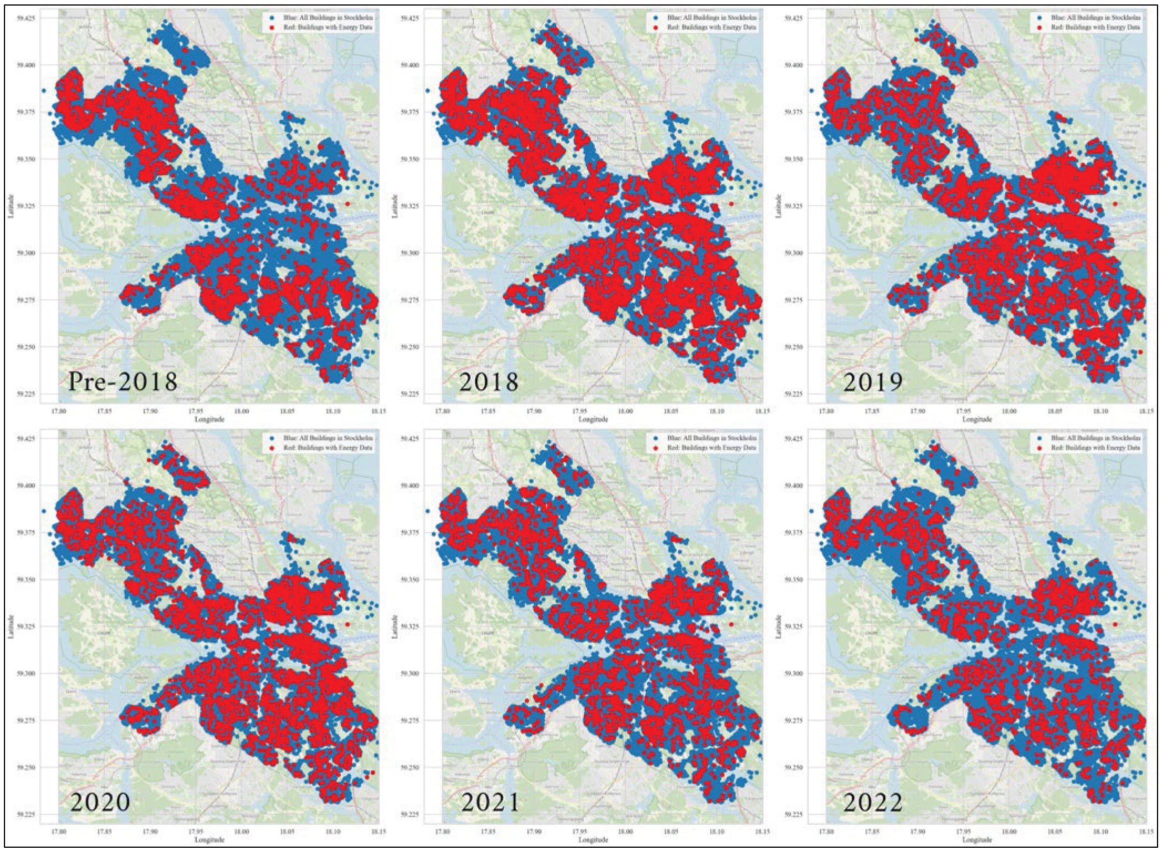

Figure 4.1 Unique buildings/properties in Stockholm (blue dots) compared with properties that have energy data (Red dots)

A heat map was created to further illustrate the spatial distribution and intensity of heat demand, as shown in Figure 4.2. This heat map visualizes the density of heat demand data points across Stockholm, using the x and y coordinates to represent longitude and latitude. Lighter, more saturated areas on the map indicate regions with greater heat demand, clearly highlighting areas where heat resources are most needed. This visualization is crucial for identifying optimal locations for future data centers, enabling a more strategic and data-driven approach to district heating planning.

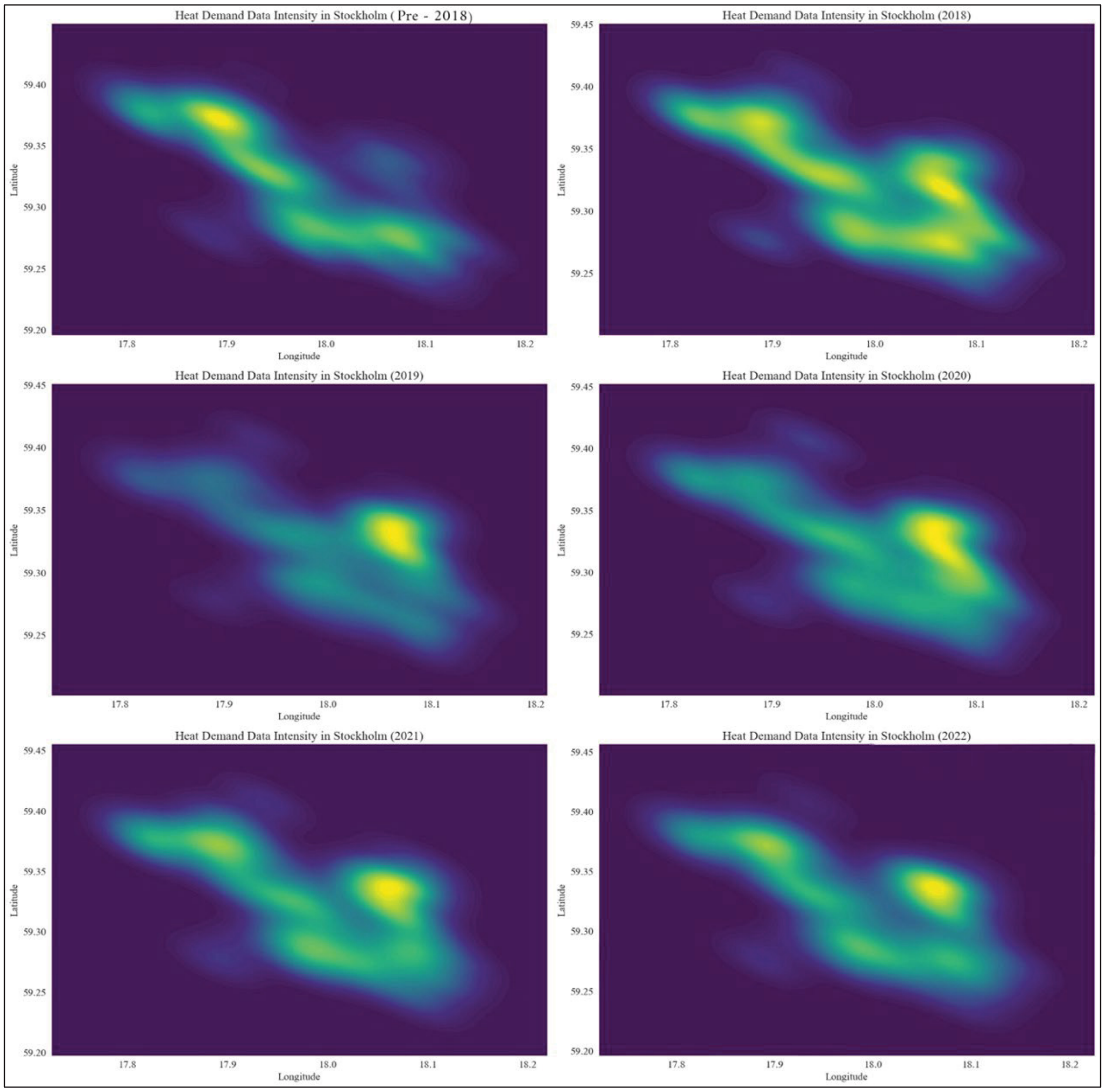

Figure 4.2 Heat demand intensity map in Stockholm ($\frac{KWh}{m^2.yr}$)

4.4.1 Allocation of future data centers using k-means and agglomerative clustering

Figure 4.3 offers a comprehensive visualization of the spatial distribution of heat demands across Stockholm, with intensity levels represented by the size and colour of the circles. Larger and darker circles indicate areas of higher heat demand, helping to identify critical zones with concentrated energy requirements. Five advanced clustering techniques K-Means, Agglomerative Clustering, DBSCAN, Spectral Clustering, and Gaussian Mixture Model

(GMM) were applied to determine the optimal locations for future data centers. Each clustering method is represented on the map, with K-Means cluster centers marked by red triangles, Agglomerative cluster centers by blue stars, DBSCAN cluster centers by blue cross, Spectral Clustering centers by purple circles, and Gaussian Mixture Model centers by green squares. This multi-method approach ensures a robust and reliable analysis, enhancing confidence in the recommended locations for data centers.

The clustering analysis reveals that Stockholm's highest heat demand concentrations are located primarily in its central and northern regions, as evident from the denser and darker circles in these areas. Each clustering method provided unique insights into the spatial distribution of heat demand. K-means and Agglomerative Clustering consistently overlapped, demonstrating their reliability in identifying high-demand zones. DBSCAN excelled in identifying smaller, denser clusters, which highlight micro-level variations in heat demand. Spectral Clustering and GMM offered additional perspectives by identifying flexible cluster shapes and providing probabilistic clusters, respectively. These nuances help refine the decision-making process and broaden the scope of infrastructure planning.

Each of the five clustering methods brings unique strengths and perspectives to the analysis, offering a comprehensive understanding of heat demand patterns in Stockholm. K-Means Clustering provides a simple and efficient solution for partitioning data into well-defined clusters, though its assumption of spherical and similarly sized clusters may not always reflect the actual spatial distribution.

Agglomerative Clustering effectively captures hierarchical relationships and yields flexible, interpretable cluster shapes, but can become computationally intensive and less scalable compared to K-Means. DBSCAN excels at identifying dense regions of heat demand, uncovering smaller clusters that may be overlooked by other methods, yet struggles with sparse data and is highly sensitive to parameter choices. Spectral Clustering accommodates complex, non-linear cluster shapes, making it ideal for intricate spatial patterns, although it is computationally expensive and sensitive to the prescribed number of clusters. Finally, the Gaussian Mixture Model (GMM) offers probabilistic clusters with nuanced boundaries that facilitate uncertainty quantification, but its performance may hinge on initial conditions and extensive parameter tuning. Together, these methods provide a robust toolkit for exploring, analyzing, and interpreting the spatial variability of heat demand in Stockholm.

No single method is universally superior. Instead, the use of multiple methods enriches the analysis by capturing different aspects of the data. For instance, DBSCAN is effective in detecting localized high-demand clusters, while GMM offers a probabilistic perspective that accounts for variations within the data. The overlaps between K-Means and Agglomerative Clustering provide consistent insights, reinforcing the reliability of their suggested locations. Spectral Clustering, with its ability to handle non-linear patterns, complements the other methods by uncovering additional potential sites.

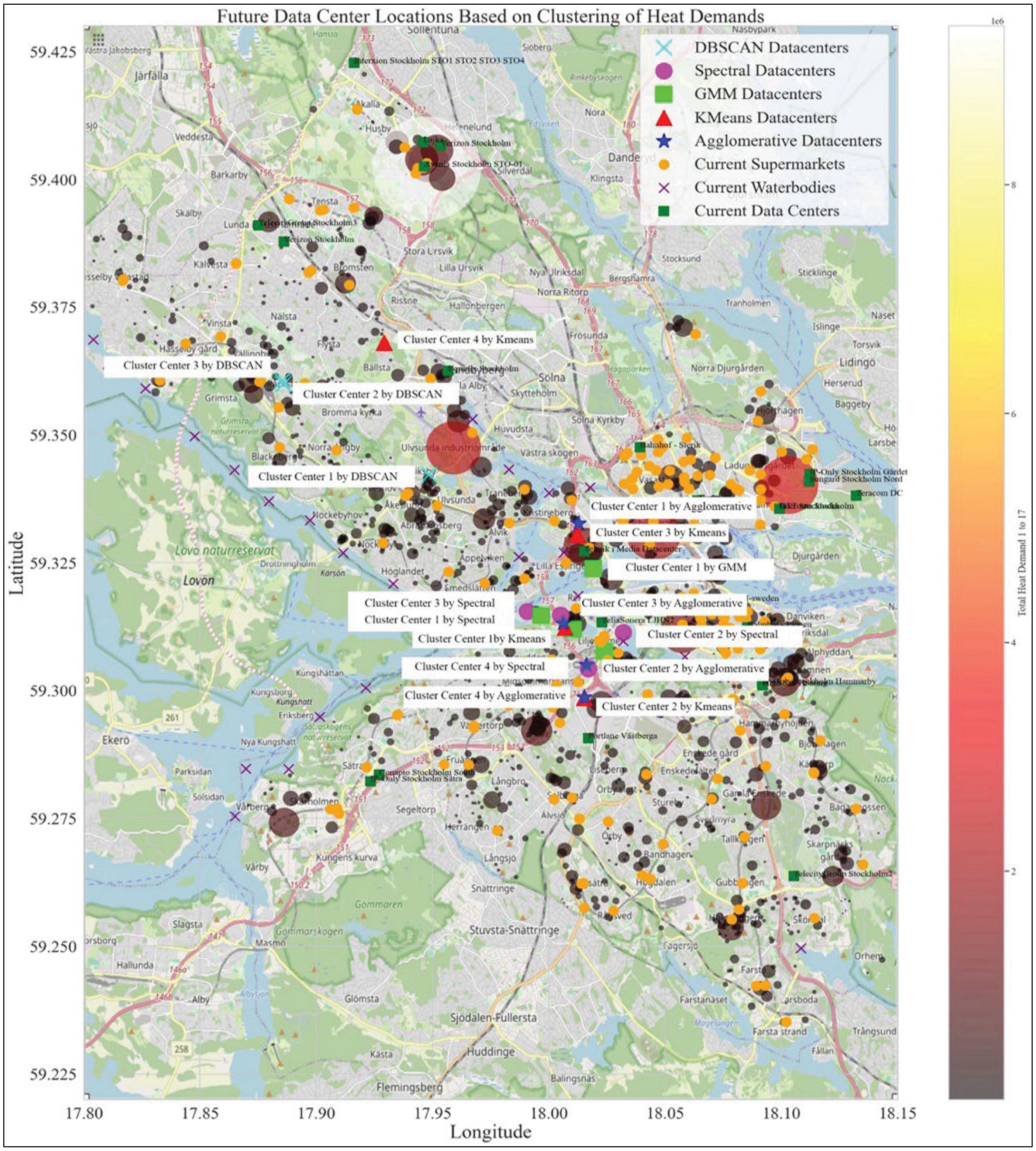

Figure 4.3 Total heat demand mapping and the locations of future data centers based on five clustering methods

Table 4.3 provides a numerical comparison of the cluster center locations identified by each method.

Table 4.3 The future data center's location in Stockholm based on five clustering methods.

Data Centers	K-means (Lat, Long)	Agglomerative (Lat, Long)	DBSCAN (Lat, Long)	Spectral (Lat, Long)	GMM (Lat, Long)
1	59.3123, 18.0070	59.3326, 18.0128	59.3422, 17.9483	59.3148, 18.0050	59.3238, 18.0191
2	59.2986, 18.0153	59.3050, 18.0162	59.3600, 17.8858	59.3042, 18.0171	59.3120, 18.0104
3	59.3305, 18.0123	59.3131, 18.0061	59.3606, 17.8853	59.3154, 17.9908	59.3076, 18.0239
4	59.3680, 17.9292	59.2986, 18.0153	59.3543, 17.9065	59.3113, 18.0321	59.3146, 17.9968

The clustering results indicate that while K-Means and Agglomerative Clustering tend to converge on similar cluster centers, DBSCAN, Spectral Clustering, and GMM identify additional or alternative locations. This divergence underscores the value of employing multiple clustering techniques to capture both global and local patterns in the data. DBSCAN highlighted micro-level heat demand variations that could be critical for localized planning. Spectral Clustering uncovered complex patterns, suggesting potential sites near existing infrastructures. GMM, with its probabilistic approach, provided insights into areas of uncertainty, enabling more informed decision-making.

The strategic placement of data centers based on these clustering methods offers significant potential for optimizing Stockholm’s district heating network. Data centers located near high-demand areas can efficiently supply heat, leveraging their excess heat for sustainable urban energy solutions. Moreover, integrating supermarkets and other existing infrastructures into the analysis further highlights opportunities for collaborative heat recovery initiatives, enhancing the overall efficiency and sustainability of the heat distribution network.

However, clustering-based approaches are inherently limited by their reliance on historical data and static assumptions. They do not account for future urban development, demographic shifts, or evolving heat demand patterns. Smaller but strategically important pockets of demand might not significantly influence the cluster centroids and could be overlooked. To address these limitations, clustering results should be complemented with predictive models, urban planning considerations, stakeholder engagement, and adaptive strategies that can accommodate future changes.

By leveraging five advanced clustering methods K-Means, Agglomerative Clustering, DBSCAN, Spectral Clustering, and GMM this analysis provides a comprehensive understanding of heat demand patterns in Stockholm.

The multi-method approach not only identifies optimal locations for future data centers but also highlights the strengths and limitations of each technique. Integrating these insights with urban planning strategies ensures a balanced, data-driven, and adaptive approach to sustainable energy solutions. This methodology demonstrates the potential for clustering techniques to guide the development of efficient district heating systems, supporting Stockholm's broader sustainability goals.

4.4.2 Allocation of heat sources to clusters

Figure 4.4.4 presents a map segmented into 10 districts using a k-means clustering algorithm accessed via the scikit learn library. Each district is assigned the heat sources within its cluster and water body sources, which are connected based on the minimum connection distance. The heat supply from each source is calculated, and the total cost is determined using the marginal cost of heat (MCOH). In Figure 4.4, each bubble represents a heat load, with the bubble size corresponding to the annual energy demand. Table 4.4 provides each cluster's total heat energy demand and the allocated residual heat supply by type. The total heating demand of 7.7 TWh/year matches the reported district heating supply of the city (Levihn et al., 2017), indicating that the EPC dataset offers comprehensive coverage.

Table 4.4 Heat demand and allocation per cluster (in GWh/year)

Cluster	Heating Demand	Data Centers	Super-markets	Water Bodies	Total Allocation
1 (red)	1,672	0	24	100	124
2 (blue)	1,006	60	26	100	186
3 (green)	464	80	18	0	98
4 (purple)	557	20	23	0	43
5 (yellow)	558	60	50	800	910
6 (pink)	959	160	104	300	564
7 (mauve)	170	40	12	500	552
8 (grey)	155	0	9	700	709
9 (mauve)	635	20	16	500	536
10 (magenta)	1548	40	17	400	457
Totals	7,754	480	299	3,400	4,179

The heating supply from environmental sources is 4.2 TWh/year, accounting for approximately 54% of the city's total annual demand. This figure contrasts with previous mapping by Su et al. (Su et al., 2021), where data centers contributed 3.2 TWh/year; in this study, their contribution is reduced to 0.48 TWh/year. This discrepancy arises from updated GIS data, which shows fewer data centers within Stockholm's administrative boundaries (the focus of this study) and indicates that the heat output from the average existing data center is lower than previously reported by Su et al., who based their findings on relatively large data centers by Nordic standards (Arizton, 2023; Wahlroos et al., 2018). Nonviable heat generation during warmer periods (12 °C and greater) is also removed here, consistent with how the Open DH market in Stockholm functions.

A distinct colour scheme is employed to differentiate the clusters on the visual map, resulting in an illustrative scatter plot that overlays normalized heat demand data onto the city's geographic layout. Each cluster is marked with unique hues, with geographic points scaled to represent their respective heat demands.

The clustering results presented in Figure 4.4 provide valuable insights into the spatial distribution of heat demand across Stockholm. By segmenting the city into distinct clusters, urban planners and policymakers can better understand the regions with the highest and most consistent heat demand, enabling targeted investments in district heating infrastructure. For example, areas with large clusters of high heat demand, such as those shown in red, yellow, and pink clusters, could be prioritized for the installation of additional heat recovery systems, including the integration of data centers, supermarkets, and water bodies as heat sources. This strategic allocation of heat sources helps to optimize energy use and minimize the need for conventional, fossil fuel based heating solutions. Furthermore, the bubble size on the map representing heat load can help policymakers allocate resources more efficiently, ensuring that areas with higher demand are adequately supplied.

In addition to these infrastructural recommendations, the heat source allocation strategy has significant economic implications. By linking data centers and supermarkets to the district heating system, cities can lower operational costs through heat recovery and reduce reliance on external energy sources. For instance, data centers in Cluster 2 (blue) can be incentivized to supply heat to the surrounding areas, especially in regions like Cluster 1 (red), where heating demand is high. This strategy could reduce the marginal cost of heat (MCOH), ensuring that heat produced from environmental sources remains cost-competitive compared to traditional heating methods. Additionally, understanding how environmental heat sources contribute to the total demand helps to shape policies that promote renewable heat generation while also balancing the economic feasibility of these systems. By incorporating such policies, local governments can align their heating strategies with broader sustainability goals, reducing carbon emissions and fostering the growth of green technologies in urban environments.

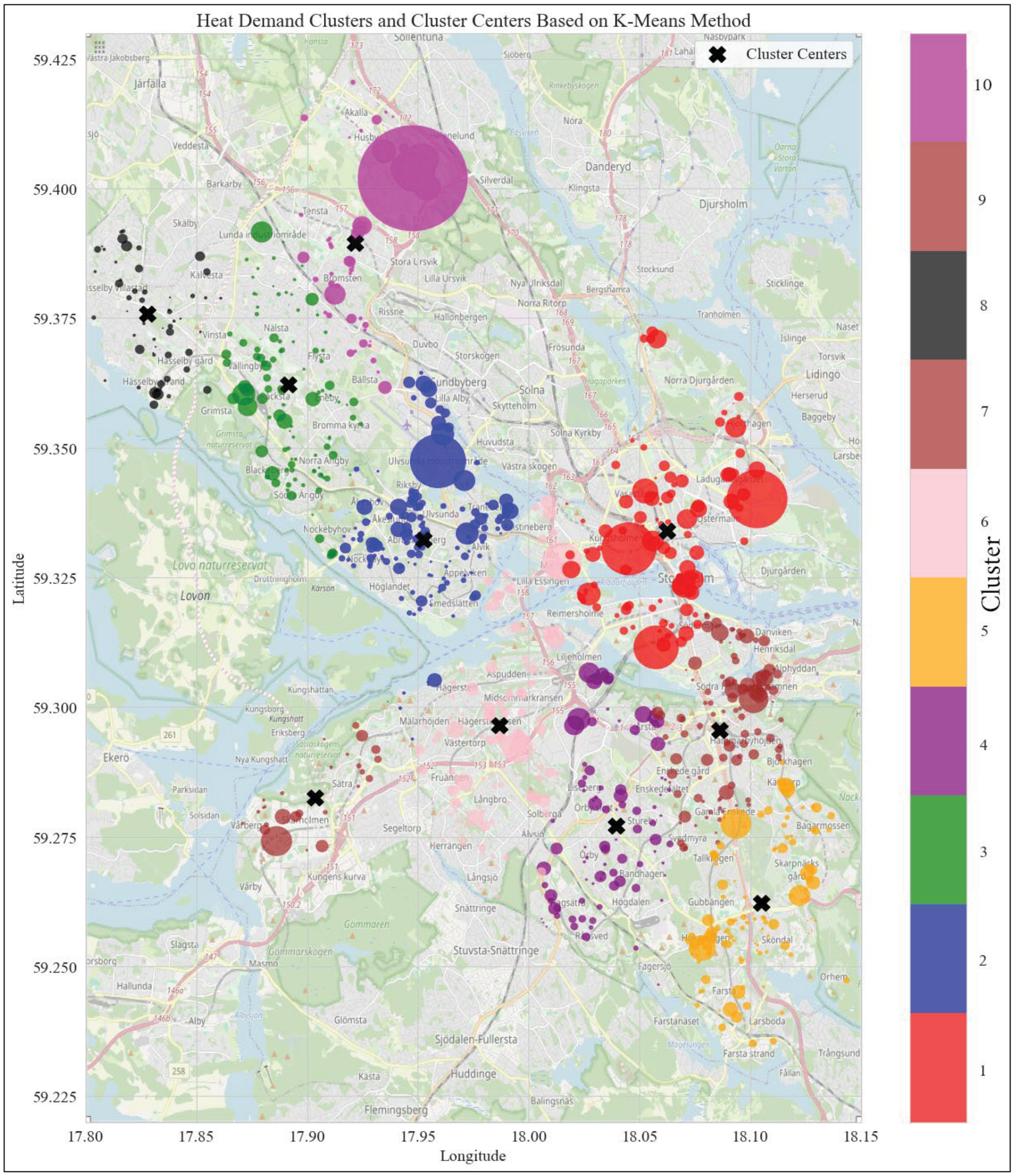

Figure 4.4 Heat demand mapping of Stockholm with 10 k-means clustered districts.

Figure 4.5 is a heatmap that shows the distances between various cluster centers, calculated based on their geographic coordinates and displayed in kilometers. The colours range from dark red (indicating smaller distances) to light red (indicating larger distances), helping to identify clusters that are close to each other and those that are farther apart.

This visualization highlights the spatial distribution of clusters, with closer centers suggesting areas of high heat demand and more distant centers indicating regions with dispersed or unpredictable demand. These insights are useful for optimizing district heating systems.

The study’s clustering results and data center allocation provide an in-depth analysis of the distance variation between different cluster centers, offering valuable insights for optimizing district heating systems. These clusters represent areas with varying heat demand across Stockholm, and clustering techniques help identify strategic locations for placing distributed heat sources, such as data centers, supermarkets, and water bodies. The variation in distances across clusters and the different clustering methods applied have significant implications for the overall system.

The centers of the clusters identified using different methods (DBSCAN, Spectral Clustering, GMM, KMeans, and Agglomerative Clustering) exhibit considerable variability. This is to be expected, as each algorithm has distinct characteristics that influence the placement of centers. DBSCAN, for example, detects denser and more localized clusters, making it ideal for pinpointing smaller hotspots of heat demand that other methods might miss. Spectral Clustering, on the other hand, identifies flexible and complex shapes, which is particularly useful for regions with irregular or non-linear heat demand patterns. The Gaussian Mixture Model (GMM) provides a probabilistic view of clusters, offering a nuanced understanding of uncertainty in heat demand, which can be beneficial for urban planners considering variability in demand patterns.

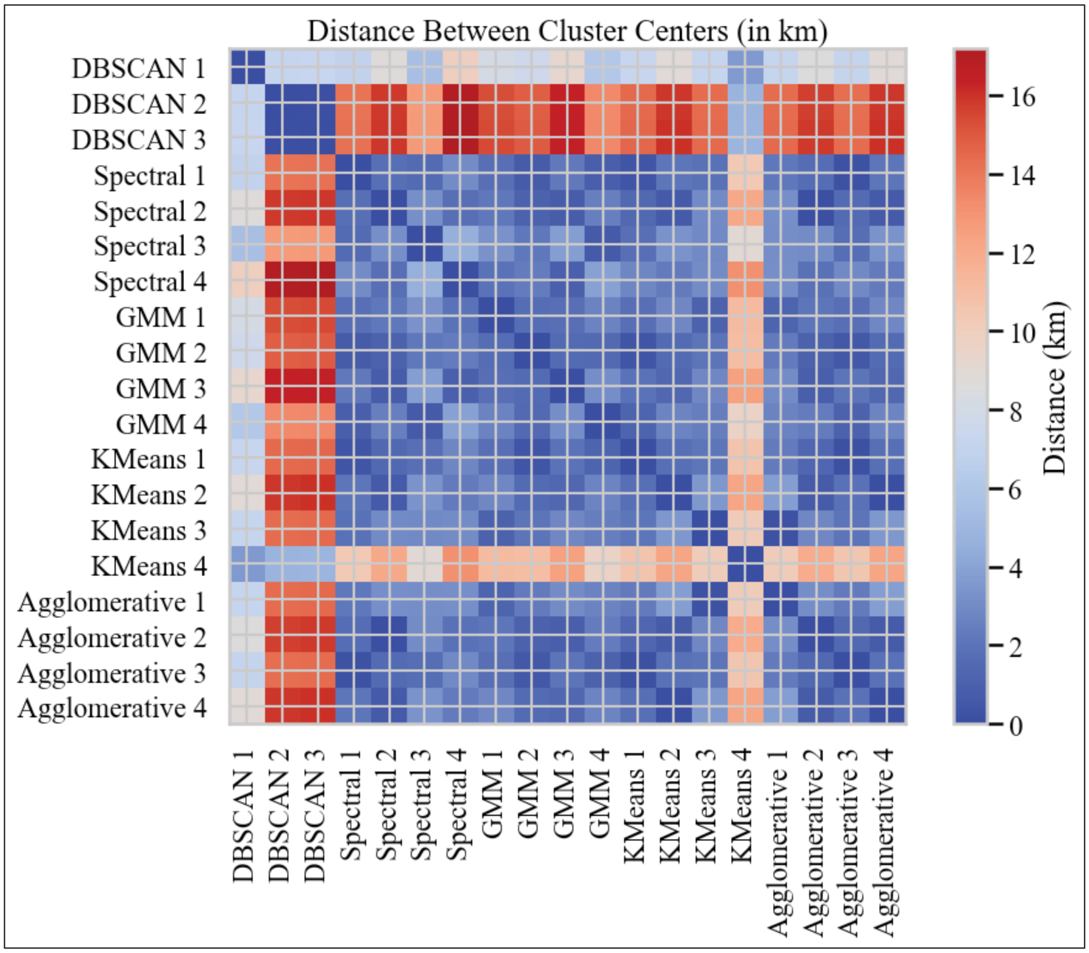

Figure 4.5 Distances between clusters and heat sources.

The distances between the cluster centers identified by each method provide important insights into the spatial distribution of heat demand. Smaller distances between centers, as seen with K-means and Agglomerative Clustering, suggest areas with well-defined high heat demand, which can be reliably targeted for infrastructure development, such as placing new data centers or optimizing existing ones for heat recovery. Larger distances, identified by methods like DBSCAN and GMM, point to regions where heat demand is more unpredictable, indicating that localized solutions, such as smaller-scale heat sources, might be needed. These methods also reveal potential gaps or underutilized areas that could benefit from heat recovery solutions. The variation in distances highlights the differing granularity of each algorithm's approach, with GMM's probabilistic nature identifying uncertain or boundary areas that could play a key role in future district heating expansions.

The variation in cluster distances directly impacts the optimization of district heating infrastructure. Clusters with tightly packed centers, as identified by methods like K-means and Agglomerative Clustering, suggest areas of high, concentrated heat demand. These regions can be prioritized for expanding or upgrading existing district heating infrastructure, facilitating the integration of heat recovery systems. Conversely, larger distances, as revealed by DBSCAN, imply more dispersed heat demand, which may benefit from decentralized heat solutions, such as localized heat pumps or smaller heat recovery units. Spectral Clustering, with its ability to capture complex clusters, may also identify areas that require tailored heat supply solutions. The economic feasibility of heat recovery is influenced by the alignment of these clusters with existing infrastructure. Areas with a higher concentration of heat demand and residual heat sources, such as data centers or water bodies, tend to have a lower marginal cost of heat (MCOH), making them more economically viable for district heating investments.

Temporal variations in heat demand, such as seasonal changes, could significantly alter the ideal cluster center locations. In colder months, heat demand is likely to be higher, necessitating the placement of heat recovery systems to meet these increased needs. During warmer months, when data centers and supermarkets generate more residual heat, clustering methods that account for dynamic heat generation, such as GMM, can provide better insights into when and where to optimize heat recovery systems. Integrating time-based data would enhance the accuracy of clustering and the strategic placement of heat recovery units, ensuring efficiency year-round.

From a strategic urban planning perspective, the clustering analysis provides crucial data for optimizing energy use and reducing reliance on fossil fuels. By identifying clusters with high heat demand and their proximity to residual heat sources, urban planners can strategically place heat generating facilities, such as data centers, in areas with high demand, minimizing infrastructure costs and ensuring that heating systems are aligned with local needs. Collaborative heat recovery initiatives, integrating sources like data centers and supermarkets into district heating networks, can reduce operational costs and carbon emissions.

The variation in cluster distances across different methods offers a comprehensive view of Stockholm's heat demand landscape. By combining these insights with considerations of economic feasibility and temporal variations, urban planners can make informed decisions that optimize both system performance and cost-effectiveness. The use of multiple clustering methods adds robustness to the analysis, ensuring that no significant heat demand areas are overlooked and that infrastructure investments align with sustainable, low-carbon goals.

4.4.3 Calculation of levelized cost of heat per cluster and city

Figure 4.6 presents a map of the marginal cost of heat (MCOH) for a range of seasonal coefficient of performances (SCOP) and electricity prices, with several key values. The zones of MCOH values (in different shades of blue) are formed by Equation 4.1, and the three solid price lines represent three different values of heat; in yellow is the weighted average sale price over the year for the Open District Heating (ODH) market (i.e. a wholesale price) at 22.0 EUR/MWh[1] , in orange is the ODH price when warm weather sales over 12 °C are removed at 27.4 EUR/MWh, and the final price in red is the 2024 seasonally weighted retail price of heat at 39.4 EUR/MWh (Exergi, 2024b).

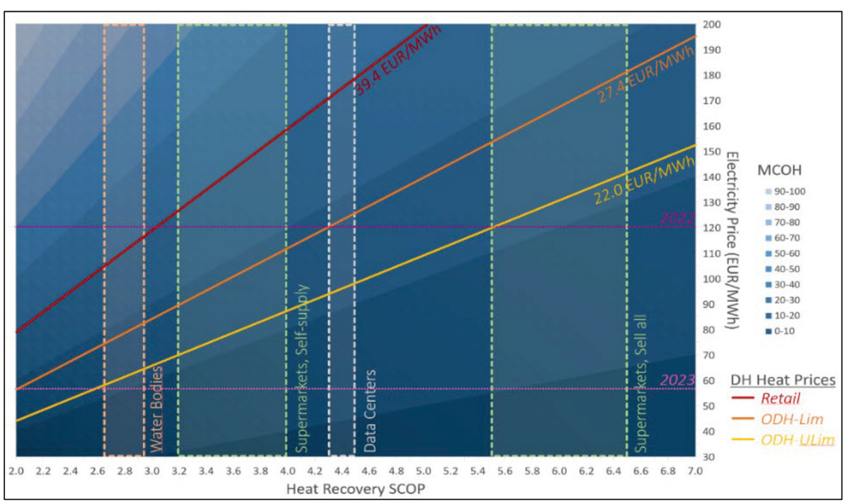

Figure 4.6 Marginal cost of heat (in EUR/MWh) by source and electricity price

[1] All prices are converted from SEK to EUR using th 2023 average conversion rate of 11.5 SEK/EUR.

Seasonal COP values for water bodies, data centers, and supermarkets are shown in Figure 4.6, with dashed boxes noting the ranges found in the literature. Water bodies are assumed to have an average SCOP of 2.8, which is relatively conservative compared to existing heat pumps in Stockholm, reaching above 3.0 (Levihn, 2017). Data centers can have many designs, including advanced water cooling, without needing a heat pump. However, the example here relies on typical air-cooled servers with an average SCOP of 4.4 (Sintong, 2023). Supermarkets can use their waste heat directly to help offset retail purchases, which offers greater value than selling to ODH and means the realized SCOP for supermarket sales can vary dramatically depending on the sales strategy (Raka Adrianto et al., 2018; Steuer D, 2023). It is assumed here that self-consumption is the preferred strategy; therefore, the SCOP for the district heating supply will be in the lower range, with an average of 3.5.

Using the electricity price and SCOPs described above, the resulting MCOH values for data centers, supermarkets, and water bodies are 12.7, 16.0, and 20.0 EUR/MWh, respectively. All prices are below the 22.0 EUR/MWh price for ODH, corroborating the results from the recent Nordic Energy Outlook (Arizton, 2023), which concluded that 98% of all residual heat sources in Stockholm are economically viable.

The total cost and weighted price for heating by cluster are given in Table 4.5, which shows a reduction in heating price in all clusters and is most significant in those with the least heat coming from the existing district heating supply. In an advanced and well-developed district heating system like Stockholm, cluster analysis may not be relevant, but the method highlights how cities without district heating may prioritize development around environmental heat sources to minimize costs.

Table 4.5 Total annual costs and weighted prices for heat supplied by residual source and cluster

Cluster	Existing DH *MEUR/yr*	Data Centers *MEUR/yr*	Super-markets *MEUR/yr*	Water Bodies *MEUR/yr*	Total Cost *MEUR/yr*	Weighted Price *EUR/MWh*
1 (red)	34.1	0.0	0.4	2.0	36.4	21.8
2 (blue)	18.0	0.8	0.4	2.0	21.2	21.1
3 (green)	8.0	1.0	0.3	0.0	9.4	20.2
4 (purple)	11.3	0.3	0.4	0.0	11.9	21.4
5 (yellow)	0.0	0.8	0.8	16.0	17.6	29.9
6 (pink)	8.7	2.0	1.7	6.0	18.4	19.2
7 (mauve)	0.0	0.5	0.2	2.4	3.1	18.0
8 (grey)	0.0	0.0	0.1	2.9	3.1	19.8
9 (mauve)	2.2	0.3	0.3	10.0	12.7	20.0
10 (magenta)	24.0	0.5	0.3	8.0	32.8	21.2
Totals	106.3	6.1	4.8	49.4	166.6	21.5

LCOH is a convenient metric for comparison, but does not capture the complexities of investment planning or operational strategies. Swedish DH networks have become largely based on boilers and/or co-heat and power (CHP) plants fueled by biomass or municipal solid waste, electric boilers, and heat pumps (Werner, 2017). CHP plants and heat pumps have complex optimization challenges, given that they interact with two independent markets: heat and electricity. Higher electricity prices favour CHP plants, which earn higher revenues, whereas heat pump supply becomes more expensive (Romanchenko et al., 2017), which is also shown in Figure 4.6. Electricity price volatility has also increased in Sweden, increasing the opportunity for arbitrage and/or ancillary services, which can moderately increase CHP revenues but are limited by the need to supply heat (Gao et al., 2022). Likewise, heat pumps can react to price volatility to minimize operational costs (Levihn, 2017).

From an investment perspective, the spatial differences in electricity and heating markets are a point of conflict for district heating supply technologies. Heating demand is local (i.e. within a city), and the electricity market is at a minimum regional (i.e. prices set by regional zones) but influenced by neighbouring price zones, often in other countries. National electricity portfolios and transmission capacity will impact the viability of heat pump and CHP investments (Beiron et al., 2022) and require deeper scenario and risk analysis than those provided in these results. If electricity price trends continue towards moderate average increases and higher volatility, then investments in thermal storage will be a key enabler of greater heat pump adoption in district heating (Holmér et al., 2020) and help maintain the cost competitiveness found in this study.

Temporal changes in heat demand, supply, and electricity prices can significantly impact clustering results and the economic outcomes of district heating systems. Heat demand fluctuates seasonally and daily, affecting the allocation of resources if not considered during clustering. Residual heat sources, such as data centers, supermarkets, and water bodies, also experience temporal variability in heat availability, with data centers' cooling demands and supermarket refrigeration systems varying across seasons. Additionally, electricity prices fluctuate, influencing the marginal cost of heat (MCOH) for these sources, making their economic viability sensitive to market conditions. Failure to incorporate temporal data in clustering could lead to suboptimal resource allocation, inefficiencies in heat supply-demand matching, and inaccurate economic evaluations. Integrating time-based data and adjusting for seasonal and hourly variations would improve the precision of clustering results and provide more sustainable, cost-effective planning for district heating networks, benefiting both urban energy planners and policymakers.

4.5 Conclusion

This research introduces a novel data-driven approach to district heating network planning by leveraging high resolution source-load mapping to address distributed heat sources' spatial and temporal variability. Unlike traditional methods that focus on centralized heat production, this study integrates detailed spatial data for distributed heat supply sources, such as data centers, supermarkets, and water bodies, alongside demand points within Stockholm. Employing advanced clustering techniques, including K-means, agglomerative clustering, DBSCAN (Density-Based Spatial Clustering of Applications with Noise), and Gaussian Mixture Model clustering, Spectral Clustering, the study robustly analyzes and identifies optimal locations for resource allocation. This multi-method approach ensures adaptability to varying data characteristics, such as noise and non-linear distributions.

Additionally, the study pioneers an in-depth techno-economic assessment of distributed heat sources, revealing that 98% of Stockholm's residual heat sources are economically viable under current market conditions. High spatial resolution (sub-kilometer scale) and temporal granularity are utilized, surpassing previous studies that rely on coarser datasets. These advancements enable precise identification of heat source locations and improved modelling of supply-demand dynamics. By focusing on Stockholm, a global leader in district heating, the research provides actionable insights and establishes a replicable methodology for other urban centers aiming to adopt sustainable heating systems.

The study’s contributions include the development of a comprehensive spatial database encapsulating detailed building energy performance metrics for Stockholm (2014–2022), enriched with extensive geographic data to support informed urban energy planning. Quantitative analysis demonstrates that environmentally friendly sources can supply 54% of Stockholm’s annual heat demand of 7.7 TWh/year, with significant contributions from data centers (0.48 TWh), supermarkets (0.3 TWh), and water bodies (3.4 TWh).

Strategic clustering methods identify optimal sites for future data centers, maximizing their role as heat sources while aligning with Stockholm’s urban energy goals. Economic analysis further confirms that distributed heat sources are cost-effective, with marginal heat costs below the current Open District Heating market price of 22.0 EUR/MWh. Specifically, the marginal costs of heat (MCOH) for data centers, supermarkets, and water bodies are estimated at 12.7 EUR/MWh, 16.0 EUR/MWh, and 20.0 EUR/MWh, respectively.

By bridging gaps in existing literature, particularly in source-load mapping resolution and the role of residual heat sources, this study underscores the importance of localized, high-resolution data. The methodology combining GIS data, unsupervised machine learning techniques, and economic analysis offers a replicable framework for optimizing district heating networks in other cities. This research not only highlights the dynamic potential of distributed heat sources but also provides strategic insights for policymakers and urban planners, setting a benchmark for sustainable and cost-effective district heating solutions.

Future studies in district heating network planning should focus on several key areas to improve and expand upon the findings of this study. First, data collection and accuracy should be enhanced, particularly by integrating high-resolution, real-time data on heat demand and supply. This would allow for dynamic modelling that accounts for fluctuations in weather, energy prices, and demand, improving system efficiency and responsiveness. Second, exploring new heat sources is essential. While this study focused on data centers, supermarkets, and water bodies, future research should investigate other renewable sources, such as geothermal energy, waste heat from industrial processes, or heat recovery from electric vehicles. This would diversify the heat supply and improve the sustainability of district heating systems. Scalability should also be a major focus. The methodology used in this study was applied to Stockholm, but future research should test the approach in other cities with varying climate conditions and energy systems. Comparative studies would help refine the models and assess their generalizability in different urban contexts.

Further, advancing clustering and machine learning techniques could provide deeper insights into heat demand patterns. Techniques like deep learning or hybrid models that combine clustering with predictive analytics could improve the accuracy of heat distribution planning. Additionally, the use of economic and policy analysis will be crucial to evaluate the long-term viability of distributed heat sources and assess the impact of different regulatory frameworks on their deployment. Future research should also explore stakeholder engagement and understand the challenges faced by building owners, utility providers, and policymakers. Social and political factors play a key role in the adoption of new technologies and should be studied to facilitate smoother integration of district heating networks. Finally, integrating district heating with other urban infrastructure, such as smart grids and waste management systems, could enhance overall energy efficiency. Studies focusing on long-term sustainability and climate adaptation are also needed to ensure that district heating networks remain resilient in the face of climate change and energy market fluctuations.

Nomenclature

Category	Term	Description
Nomenclatures and Abbreviations	EPC	Energy Performance Certificates
	GIS	Geographic Information Systems
	ODH	Open District Heating
	TWh	Terawatt-hour
	COP	Coefficient of Performance
	MCOH	Marginal Cost of Heat
	SCOP	Seasonal Coefficient of Performance
	DH	District Heating
	DBSCAN	Density-Based Spatial Clustering of Applications with Noise
	GMM	Gaussian Mixture Model
	KMeans	K-means clustering algorithm
	HAC	Hierarchical Agglomerative Clustering
	UEUs	Urban Energy Units
	RHI	Renewable Heat Incentive
Parameters and Variables	n_clusters	Number of clusters for clustering algorithms
	eps	Maximum distance between two samples for DBSCAN
	min_samples	Minimum number of points required to form a "core" point in DBSCAN
	affinity	Metric used to compute similarity in Spectral Clustering
	reg_covar	Regularization for ensuring covariance matrices are positive semi-definite in GMM
	max_iter	Maximum number of iterations for GMM
	imputer_strategy	Strategy for imputing missing values (mean strategy)

Category	Term	Description
Parameters and Variables	scaler	Standardization method (StandardScaler)
	Qj,i	Heat supply for each source in the cluster
	Qi	Total heat demand of the cluster
	KMeans Centroids	Centroids of clusters identified by KMeans
	Ward's Method	Method used in Agglomerative Clustering to minimize the variance within a cluster
	Cluster Center Locations	Locations of cluster centers as identified by different clustering methods

CONCLUSION

The conclusion of this study reflects the critical considerations in utilizing Machine learning and Deep learning algorithms for predicting, evaluating, and recommending energy load measures and analysis in district areas. Based on fundamental physical principles, white box models provide deep and interpretable insights into the underlying mechanisms of energy systems, making them ideal for detailed analysis, retrofitting, and compliance verification. However, they are often time-intensive and complex, limiting their applicability for large-scale or real-time tasks. In contrast, black box models powered by machine learning and statistical methods offer the advantage of speed, scalability, and predictive accuracy without requiring a comprehensive understanding of the internal system dynamics. For this reason, black box models were selected for this analysis, particularly for optimizing operations and real-time energy management, where rapid and accurate predictions are essential. By leveraging machine learning, the models were able to capture intricate patterns in energy consumption data, ensuring better results in forecasting tasks across multiple time frames.

The review of building energy management simulation tools reveals a diverse array of tools designed to meet the complex challenges of energy efficiency optimization in buildings. White box models, grounded in fundamental physics, offer detailed and accurate analysis, making them suitable for tasks like energy performance assessments, retrofitting projects, and compliance checks. Conversely, black box models are better suited for quick, operational decision-making and real-time management, allowing researchers and consultants to generate fast, actionable results. Additionally, including web-based simulation tools further enriches the landscape, providing accessible platforms for data visualization and benchmarking. The choice between these modelling approaches depends largely on the specific goals and complexity of the project. In cases of large and complex buildings, scalable solutions are necessary to strike a balance between accuracy and efficiency.

Ultimately, this review emphasizes the importance of a flexible, tailored approach, enabling professionals to make data-driven decisions that optimize energy use and enhance sustainability in the built environment. Integrating these tools allows for better collaboration between data scientists, engineers, and analysts, facilitating informed decision-making that promotes energy efficiency and sustainable practices. As the field evolves, building energy management simulation tools remain essential in helping professionals meet energy challenges, innovate in building design, and drive advancements in sustainable operations.

The comprehensive analysis of energy consumption patterns at the UBC campus in Vancouver reveals intricate relationships between various energy sources, environmental factors, and operational strategies. Using machine learning models and a cost correlation analysis, significant interdependencies were identified among electricity, hot water power, gas volume, and steam volume consumption. These findings highlight the importance of an integrated energy management approach. A particularly notable finding is the inverse correlation between Vancouver's temperature and multiple energy metrics, suggesting that energy efficiency strategies that respond to temperature variations, particularly during warmer periods, could yield significant improvements.

This study's evaluation of multiple regression models revealed that advanced machine learning techniques, especially deep neural networks, were superior in capturing complex, non-linear energy consumption patterns in the district areas. These models outperformed traditional approaches in terms of predictive accuracy, offering valuable, data-driven insights that can be used to optimize energy consumption, reduce waste, and lower environmental impact. The success of deep learning models in this context aligns with broader trends in energy management research, where advanced analytics and machine learning techniques are increasingly recognized for their ability to handle dynamic and multi source energy systems. This research offers substantial implications for energy management strategies at the UBC campus and other district areas, demonstrating that understanding the complex interplay between energy sources and environmental variables can enable more targeted and effective interventions.

Furthermore, the findings support the broader theoretical understanding that advanced machine learning algorithms hold significant promise for improving energy efficiency and operational performance in complex energy systems. Future research could further investigate the causal relationships between energy consumption and external factors, assess the specific impacts of operational interventions, and explore renewable energy integration within the campus energy portfolio. By continuing to build upon the methodologies developed in this study, institutions like UBC can lead the way in sustainable energy management while contributing to the broader body of knowledge on energy systems optimization.

By leveraging the experience of selecting black box and machine learning models and evaluating their performance in predicting district energy demand, methods for Source Load Mapping in Distributed District Heating Planning were chosen. This extensive study on Stockholm's district heating potential highlights the significant opportunities for integrating distributed heat sources into urban energy systems. District heating (DH) networks are centralized systems that distribute heat from a central source to multiple buildings, efficiently supplying heating and hot water. These systems can utilize various heat sources, including waste heat from industrial processes, data centers, supermarkets, and natural water bodies. Integrating waste heat into DH networks is particularly important because it recycles excess energy that would otherwise be lost, reducing the need for additional fuel consumption. Data centers, for example, generate substantial heat during operation, which can be captured and fed into the DH network. Similarly, supermarkets produce waste heat from refrigeration systems, and water bodies can serve as renewable heat sources through heat pumps. This approach improves energy efficiency and helps reduce greenhouse gas emissions, contributing to a more sustainable urban energy system.

By applying high-resolution spatial and temporal data analysis, optimal locations for future data centers that could supply waste heat to the district heating (DH) network were identified. This integration helps improve the efficiency of the DH system by utilizing excess heat from data centers as a valuable energy source for heating buildings. The results demonstrate that residual heat sources such as those from data centers, supermarkets, and water bodies could meet approximately 54% of Stockholm’s annual heating demand. This highlights residual heat's vast, untapped potential as a key component of the city's energy strategy.

Moreover, the economic analysis showed that many of these residual heat sources are viable and cost-effective, with marginal heat generation costs below current market prices for open district heating. This suggests that integrating these sources could lead to more sustainable and affordable heating solutions for the city.

Using a clustering approach to analyze the spatial distribution of heat sources provides critical insights for urban planning and infrastructure development. However, other methods should complement this approach to ensure a comprehensive and adaptive strategy that addresses the evolving needs of urban energy systems. The discrepancies between this study's findings and previous research, particularly regarding the contributions of data centers, further underscore the importance of using up-to-date and localized data when conducting such analyses. This research contributes to the growing body of knowledge on sustainable urban energy systems and offers a replicable methodology for other cities looking to optimize their district heating networks. Future studies could explore the dynamic patterns of urban development and the potential for integrating emerging technologies in heat recovery and distribution, further enhancing the sustainability of urban energy systems.

The original contribution of this thesis lies in developing a data-driven framework for urban and district energy load forecasting and optimization, employing advanced machine learning and deep learning techniques to tackle the complex nature of energy demand in urban environments. This framework is applied in real-world scenarios, including Stockholm's district heating systems and the UBC campus in Vancouver, demonstrating the adaptability and scalability of these predictive models across different urban settings. The research significantly advances urban energy management practices by utilizing non-linear predictive models that account for variables like weather patterns, occupancy trends, and environmental conditions.

Moreover, the thesis establishes a robust evaluation methodology to balance predictive accuracy, interpretability, and computational efficiency, selecting optimal models for realworld applications. Integrating diverse data sources, this framework improves load forecasting and provides actionable recommendations for sustainable urban planning and efficient energy distribution. These contributions offer a replicable methodology for cities aiming to enhance sustainability and resilience in their energy systems, supporting broader smart city initiatives and sustainable infrastructure development.

Overall, this project illustrates the importance of advanced analytical methods in energy systems research. The conclusions emphasize the value of tailored approaches, whether through black box models for real-time operational optimization, comprehensive correlation analyses for campus-wide energy strategies, or high-resolution spatial studies for urban heating systems. Leveraging modern data-driven techniques contributes to the growing work on optimizing energy systems for efficiency, sustainability, and cost-effectiveness. These findings have broad implications for both academic research and practical applications in energy management, highlighting the potential for continued innovation in the field.

BIBLIOGRAPHY

A.Bhatia. Cooling Load Calculations and Principles. https://www.cedengineering.com/userfiles/Cooling%20Load%20Calculations%20and%20Principles%20R1.pdf

AB, E. S. IDA Indoor Climate and Energy. Retrieved 1 Sep,2023 from https://www.equa.se/en/ida-ice

AG, N. NEPLAN. Retrieved 1 Sep,2023 from https://www.neplan.ch/en-company/

Agency, S. E. (2023). Energy in Sweden - Facts and Figures 2023. https://www.energimyndigheten.se/en/news/2023/energy-in-sweden---facts-and-figures-2023/

Agupugo, C. P., Ajayi, A. O., Nwanevu, C., & Oladipo, S. S. (2022). Policy and regulatory framework supporting renewable energy microgrids and energy storage systems. Eng. Sci. Technol. J, 5, 2589-2615.

Ahmad, T., & Chen, H. (2019). Nonlinear autoregressive and random forest approaches to forecasting electricity load for utility energy management systems. Sustainable Cities and Society, 45, 460-473.

Amara, F., Agbossou, K., Cardenas, A., Dubé, Y., & Kelouwani, S. (Year). Comparison and simulation of building thermal models for effective energy management. University of Quebec at Trois-Rivières.

Akguc, A., Gali, G., & Yilmaz, A. Z. (2013). Including the building energy performance consultancy to the integrated building design process: The industrial building case study in Turkey. VII. CLIMAMED Mediterranean Congress of Climatization,

al, J. d. B. e. (2024a). sklearn.cluster.AgglomerativeClustering. Retrieved April 16th from https://scikit-learn.org/stable/modules/generated/sklearn.cluster.AgglomerativeClustering.html

al, J. d. B. e. (2024b). sklearn.cluster.KMeans. Retrieved April 16th from https://scikit-learn.org/stable/modules/generated/sklearn.cluster.KMeans.html

Allegrini, J., Orehounig, K., Mavromatidis, G., Ruesch, F., Dorer, V., & Evins, R. (2015). A review of modelling approaches and tools for the simulation of district-scale energy systems. Renewable and Sustainable Energy Reviews, 52, 1391-1404.

Amber, K. P., Aslam, M. W., Mahmood, A., Kousar, A., Younis, M. Y., Akbar, B., Chaudhary, G. Q., & Hussain, S. K. (2017). Energy consumption forecasting for university sector buildings. Energies, 10(10), 1579.

Ang, Y. Q., Berzolla, Z. M., Letellier-Duchesne, S., Jusiega, V., & Reinhart, C. (2022). UBEM. io: A web-based framework to rapidly generate urban building energy models for carbon reduction technology pathways. Sustainable Cities and Society, 77, 103534.

Antal, M., Cioara, T., Anghel, I., Gorzenski, R., Januszewski, R., Oleksiak, A., Piatek, W., Pop, C., Salomie, I., & Szeliga, W. (2019). Reuse of data center waste heat in nearby neighborhoods: A neural networks-based prediction model. Energies, 12(5), 814.

Arendt, K., Jradi, M., Shaker, H. R., & Veje, C. (2018). Comparative analysis of white-, gray- and black box models for thermal simulation of indoor environment: Teaching building case study. Proceedings of the 2018 Building Performance Modeling Conference and SimBuild co-organized by ASHRAE and IBPSA-USA, Chicago, IL, USA,

Arizton. (2023). Arizton. Nordic Data Center Construction Market - Industry Outlook & Forecast 2023-2028. 2023.

Associates, E. a. J. J. H. eQuest the QUick Energy Simulation Tool. Retrieved 1 Oct,2023 from https://www.doe2.com/equest/

AutoDesk. Autodesk Insight 360. https://insight360.autodesk.com/oneenergy

AutoDesk. (2015). Green Building Studio. Retrieved 1 Sep,2023 from https://gbs.autodesk.com/gbs

Baetens, R., De Coninck, R., Jorissen, F., Picard, D., Helsen, L., & Saelens, D. (2015). Openideas-an open framework for integrated district energy simulations,Leuven, Belgium. Building simulation,

Baniyounes, A. M., Ghadi, Y. Y., & Baker, A. A. (2019). Institutional smart buildings energy audit. International Journal of Electrical & Computer Engineering (2088-8708), 9(2).

Beiron, J., Göransson, L., Normann, F., & Johnsson, F. (2022). A multiple system level modeling approach to coupled energy markets: Incentives for combined heat and power generation at the plant, city and regional energy system levels. Energy, 254, 124337.

Blanco Bohorquez, L. A., Alhamwi, A., Schiricke, B., & Hoffschmidt, B. (2023). Data-driven classification of Urban Energy Units for district-level heating and electricity demand analysis. Sustainable Cities and Society(101).

Blonsky, M., Maguire, J., McKenna, K., Cutler, D., Balamurugan, S. P., & Jin, X. (2021). OCHRE: The object-oriented, controllable, high-resolution residential energy model for dynamic integration studies. Applied Energy, 290, 116732.

Borioli, E., Ciapessoni, E., Cirio, D., & Gaglioti, E. (2009). Applications of neural networks and decision trees to energy management system functions. 2009 15th International Conference on Intelligent System Applications to Power Systems,

Boverket – the Swedish National Board of Housing, Building and Planning. (2024). Retrieved April 16th from https://www.boverket.se/en/start/

Brownlee, J. (2016). Machine learning mastery with Python: understand your data, create accurate models, and work projects end-to-end. Machine Learning Mastery.

Brownlee, J. (2020). How to Decompose Time Series Data into Trend and Seasonality

Retrieved April 25th from https://machinelearningmastery.com/decompose-time-series-data-trend-seasonality/

Building Energy Platform,marcelo bastos. In. (2014). https://github.com/buildingenergy/buildingenergy-platform

Building Performance Database (BPD). Retrieved 1 sep,2023 from https://www.energy.gov/eere/buildings/building-performance-database-bpd

Buildsim. Buildsim. Retrieved 1 sep,2023 from https://www.buildsim.io

Calise, F., & Figaj, R. (2022). Recent Advances in Sustainable Energy and Environmental Development. Energies, 15(18), 6534.

Canada, G. o. RetScreen. Retrieved 1 Sep,2023 from https://www.nrcan.gc.ca/maps-tools-and-publications/tools/modelling-tools/retscreen/7465

Chen, Y., Guo, M., Chen, Z., Chen, Z., & Ji, Y. (2022). Physical energy and data-driven models in building energy prediction: A review. Energy Reports, 8, 2656-2671.

Chen, Y., Shi, Y., & Zhang, B. (2017). Modeling and optimization of complex building energy systems with deep neural networks. 2017 51st Asilomar Conference on Signals, Systems, and Computers,

Chou, J.-S., & Tran, D.-S. (2018). Forecasting energy consumption time series using machine learning techniques based on usage patterns of residential householders. Energy, 165, 709-726.

Cichowicz, R., & Jerominko, T. (2023). Comparison of calculation and consumption methods for determining Energy Performance Certificates (EPC) in the case of multi-family residential buildings in Poland (Central-Eastern Europe). Energy, 282, 128393.

Ciulla, G., & D'Amico, A. (2019). Building energy performance forecasting: A multiple linear regression approach. Applied energy, 253, 113500.

Claudio, R. (2020). VZWAM Web-Based Lookup.

ClimaPlus. Retrieved 1 sep,2023 from http://climaplusbeta.com

Columbia, T. U. o. B. (Sep 2023). UBC Adress map. Retrieved June 21 from https://planning.ubc.ca/about-us/campus-maps

Crawley, D. B., Hand, J. W., Kummert, M., & Griffith, B. T. (2008). Contrasting the capabilities of building energy performance simulation programs. Building and environment, 43(4), 661-673.

Danish Building Research Institute, B. (2015). https://build.dk/Pages/SBi-Not-Found.aspx?requestUrl=https://build.dk/en/bsim

Davies, G., Maidment, G., & Tozer, R. (2016). Using data centres for combined heating and cooling: An investigation for London. Applied Thermal Engineering, 94, 296-304.

de Wilde, P., & Augenbroe, G. (2018). Energy modelling. In A Handbook of Sustainable Building Design and Engineering (pp. 95-108). Routledge.

Deb, C., & Schlueter, A. (2021). Review of data-driven energy modelling techniques for building retrofit. Renewable and Sustainable Energy Reviews, 144, 110990.

DEKSoft. Retrieved 1 sep,2023 from https://deksoft.eu/en/programy/programs

Delzendeh, E., Wu, S., Lee, A., & Zhou, Y. (2017). The impact of occupants' behaviours on building energy analysis: A research review. Renewable and Sustainable Energy Reviews, 80, 1061-1071.

Digitemie, W. N., & Ekemezie, I. O. (2024). A comprehensive review of Building Energy Management Systems (BEMS) for improved efficiency. World Journal of Advanced Research and Reviews, 21(3), 829-841.

Ding, Y., & Liu, X. (2020). A comparative analysis of data-driven methods in building energy benchmarking. Energy and Buildings, 209, 109711.

Dissanayake, N., & Dias, K. (2017). Web-based applications: extending the general perspective of the service of web. 10th International Research Conference of KDU (KDU-IRC 2017) on Changing Dynamics in the Global Environment: Challenges and Opportunities. Rathmalana,

Doe, J. S., A. (2023). Advanced Simulation Tools for Energy Efficiency. Int. J. Sustain. Des.

Dong, W., Sun, H., Li, Z., & Yang, H. (2024). Design and optimal scheduling of forecasting-based campus multi-energy complementary energy system. Energy, 133088.

Drury B.Crawley, L. K. L., Frederick C.Winkelmann, W.F.Buhlc, Y.JoeHuang, Curtis O.Pedersen , Richard K.Strand, Richard J.Liesen, Daniel E.Fisher, Michael J.Witte, Jason Glazer. (2001). EnergyPlus: creating a new-generation building energy simulation program. Energy and Buildings. https://doi.org/https://doi.org/10.1016/S0378-7788(00)00114-6

Edge App. Retrieved 1 Sep,2023 from https://app.edgebuildings.com/user/welcome?_ga=2.135834418.206550790.1587835695-1849169681.1587835695

Energinet Energy Management Software. Retrieved 1 sep,2023 from http://cebyc.com

EnerGis. EnerGis. Retrieved 1 Sep,2023 from https://www.energis.cloud/en/

Energy, R. DISTRICT ENERGY IN CITIES.

Energy, U. d. o. Un department of Energy. Retrieved 1 Aug 2023 from https://www.energy.gov/eere/buildings/downloads/energyplus-0

Energy, U. S. D. o. (2015). Energy efficiency and renewable energy, Ener-

gyPlus. Retrieved 1 Sep,2023 from https://energyplus.net

Energy., U. D. o. Retrieved 1 August from https://www.energy.gov/eere/buildings/articles/energyplus

Energyplus. Energyplus. Retrieved 1 Sep,2023 from https://energyplus.net/documentation

EnergySoft. Energysoft. Retrieved 1 Sep,2023 from http://www.energysoft.com/faqs/

EnerLogic and James J. Hirsch & Associates, DOE2,. (2012). Retrieved 1 Oct,2023 from https://www.doe2.com

EnerPro. Retrieved 1 Oct,2023 from http://www.energyprofiletool.com/subscription/default.asp

EnExPlan. Retrieved 1 Sep,2023 from http://www.almirantacorporation.com

EPWMAP. Retrieved 1 sep,2023 from https://www.ladybug.tools/epwmap/

Exergi, S. (2024a). Heat Recovery. Retrieved April 11 from https://www.stockholmexergi.se/en/heat-recovery/

Exergi, S. (2024b). Prices for condominium associations

Retrieved June 6th from https://www.stockholmexergi.se/bostadsrattsforening/vadkostardetbostadsrattsforening/

Fan, C., Xiao, F., & Zhao, Y. (2017). A short-term building cooling load prediction method using deep learning algorithms. Applied Energy, 195, 222-233.

Ferrando, M., Causone, F., Hong, T., & Chen, Y. (2020). Urban building energy modeling (UBEM) tools: A state-of-the-art review of bottom-up physics-based approaches. Sustainable Cities and Society, 62, 102408.

Forouzandeh, N., Tahsildoost, M., & Zomorodian, Z. S. (2021). A review of web-based building energy analysis applications. Journal of Cleaner Production, 306, 127251.

Foucquier, A., Robert, S., Suard, F., Stéphan, L., & Jay, A. (2013). State of the art in building modelling and energy performances prediction: A review. Renewable and Sustainable Energy Reviews, 23, 272-288.

Fritz, R. M. M., A. Radiance. Retrieved 1 Aug,2023 from https://www.radiance-online.org//

Gao, S., Jurasz, J., Li, H., Corsetti, E., & Yan, J. (2022). Potential benefits from participating in day-ahead and regulation markets for CHPs. Applied Energy, 306, 117974.

García-Fuentes, M., Hernández, G., Serna, V., Martín, S., Álvarez, S., Lilis, G., Giannakis, G., Katsigarakis, K., Mabe, L., & Oregi, X. (2019). OptEEmAL: Decision-Support Tool for the Design of Energy Retrofitting Projects at District Level. IOP Conference Series: Earth and Environmental Science, Gassar, A. A. A., & Cha, S. H. (2020). Energy prediction techniques for large-scale buildings towards a sustainable built environment: A review. Energy and Buildings, 224, 110238.

Gelažanskas, L., & Gamage, K. A. (2015). Forecasting hot water consumption in residential houses. Energies, 8(11), 12702-12717.

GEnergy. Retrieved 1 Sep,2023 from https://greenspacelive.com/site/products/genergy/

Géron, A. (2022). Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow. " O'Reilly Media, Inc.".

Giunta, F., & Sawalha, S. (2021). Techno-economic analysis of heat recovery from supermarket's CO2 refrigeration systems to district heating networks. Applied Thermal Engineering, 193, 117000.

Granlund. RIUSKA. Retrieved 1 Sep,2023 from https://www.granlund.fi/en/software/riuska/

Grove, R. F. (2009). Web Based Application Development. Jones & Bartlett Publishers.

Gupta, S., & Gupta, B. B. (2017). Detection, avoidance, and attack pattern mechanisms in modern web application vulnerabilities: present and future challenges. International Journal of Cloud Applications and Computing (IJCAC), 7(3), 1-43.

Hadziomerovic, D. (2019). Energy Systems Simulation on an Urban District-Level. In.

Hamdaoui, M.-A., Benzaama, M.-H., El Mendili, Y., & Chateigner, D. (2021). A review on physical and data-driven modeling of buildings hygrothermal behavior: Models, approaches and simulation tools. Energy and Buildings, 251, 111343.

Hamilton, I., Rapf, O., Kockat, D. J., Zuhaib, D. S., Abergel, T., Oppermann, M., Otto, M., Loran, S., Fagotto, I., & Steurer, N. (2020). 2020 global status report for buildings and construction. United Nations Environmental Programme.

Han, B., Zhang, S., Qin, L., Wang, X., Liu, Y., & Li, Z. (2022). Comparison of support vector machine, Gaussian process regression and decision tree models for energy consumption prediction of campus buildings. 2022 8th International Conference on Hydraulic and Civil Engineering: Deep Space Intelligent Development and Utilization Forum (ICHCE),

Harish, V., & Kumar, A. (2016). A review on modeling and simulation of building energy systems. Renewable and Sustainable Energy Reviews, 56, 1272-1292.

Harris, D. (2016). A guide to energy management in buildings. Routledge.

Helsinki, C. o. (2024). Energy and Climate Atlas. Retrieved April 11 from https://kartta.hel.fi/3d/atlas/#/

Hernández, G., Serna, V., & García-Fuentes, M. Á. (2017). Design of energy efficiency retrofitting projects for districts based on performance optimization of District Performance Indicators calculated through simulation models. Energy Procedia, 122, 721-726.

Hiller, M., Holst, S., Knirsch, A., & Schuler, M. (2001). TRNSYS 15-A simulation tool for innovative concepts. Proceedings of the Building Simulation'01 Conference,

HippoCMMS. Retrieved 1 Sep,2023 from https://hippocmms.iofficecorp.com

Historical Climate Data by Government of Canada (https://climate.weather.gc.ca

Hokamp, C. M. (2018). Deep interactive text prediction and quality estimation in translation interfaces Dublin City University].

Holmér, P., Ullmark, J., Göransson, L., Walter, V., & Johnsson, F. (2020). Impacts of thermal energy storage on the management of variable demand and production in electricity and district heating systems: a Swedish case study. International Journal of Sustainable Energy, 39(5), 446-464.

Home Energy Saver. Retrieved 1 Oct,2023 from https://hes.lbl.gov/consumer/

Home Energy Score. Retrieved 1 Oct,2023 from https://www.energy.gov/eere/buildings/articles/home-energy-score

Hong, T., Chen, Y., Lee, S. H., & Piette, M. A. (2016). CityBES: A web-based platform to support city-scale building energy efficiency. Urban Computing, 14, 2016.

Hong, T., Chen, Y., Luo, X., Luo, N., & Lee, S. H. (2020). Ten questions on urban building energy modeling. Building and environment, 168, 106508.

Hopfe, C. J., McLeod, R. S., & Rollason, T. (2017). Opening the black box: Enhancing community design and decision making processes with building performance simulation. The 15th International Conference of IBPSA,

Huang, M. N., T. (2023). Best Practices in Energy Consulting. Energy Consult, 32, 17.

Huang, P., Copertaro, B., Zhang, X., Shen, J., Löfgren, I., Rönnelid, M., Fahlen, J., Andersson, D., & Svanfeldt, M. (2020). A review of data centers as prosumers in district energy systems: Renewable energy integration and waste heat reuse for district heating. Applied Energy, 258, 114109.

Hung, N. T. (2020). Data-driven predictive models for daily electricity consumption of academic buildings. Aims Energy, 8(5).

IBPSA-USA. Building Energy Software Tools. https://www.buildingenergysoftwaretools.com

Idowu, S., Åhlund, C., & Schelen, O. (2014). Machine learning in district heating system energy optimization. 2014 IEEE International Conference on Pervasive Computing and Communication Workshops (PERCOM WORKSHOPS),

IES. Integrated Environmental Solutions Ltd., IESVE. Retrieved 1 Sep,2023 from https://www.iesve.com

IFAF. (2024). Visualisierung von Heizenergieverschwendungen in ¨offentlichen Geb¨auden durch eine Heatmap. https://www.ifaf-berlin.de/projekte/heatmap/

Imbert, C., Bhattacharjee, S., & Tencar, J. (2018). Simulation of urban microclimate with SOLENE-microclimat An outdoor comfort case study. Simul. Ser, 50, 198-205.

Institute, D. B. R. (2015). Retrieved 1 Oct,2023 from https://build.dk/Pages/SBi-Not-Found.aspx?requestUrl=https://build.dk/miljo-og-energi/energiberegning

Institute for Energy efficient Buildings and indoor climate , R. A. U. PyCity. In https://github.com/RWTH-EBC/pyCity

Intelligent Environments Laboratory,CityLearn. In. https://github.com/intelligent-environments-lab/CityLearn

Ismanto, R. N., & Salman, M. (2017). Improving security level through obfuscation technique for source code protection using AES algorithm. Proceedings of the 2017 the 7th International Conference on Communication and Network Security,

Jeon, G. (2022). Artificial Intelligence Approaches for Energies. In (Vol. 15, pp. 6651): MDPI.

Korpela, T., Kuosa, M., Sarvelainen, H., Tuliniemi, E., Kiviranta, P., Tallinen, K., & Koponen, H.-K. (2021). Waste heat recovery potential in residential apartment buildings in Finland's Kymenlaakso region by using mechanical exhaust air ventilation and heat pumps.

Kräuchi, P., Kolb, M., Gautschi, T., Menti, U.-P., & Sulzer, M. (2014). Modellbildung für thermische Arealvernetzung mit IDA-ICE. Fifth German-Austrian IBPSA Conference-RWTH Aachen University,

KTH University Stockholm (https://www.liveinlab.kth.se/en/start-1.1064463

Laboratory, N. R. E. Open Studio. Retrieved 1 Sep,2023 from https://openstudio.net

Lee, S. H., Hong, T., Piette, M. A., & Taylor-Lange, S. C. (2015). Energy retrofit analysis toolkits for commercial buildings: A review. Energy, 89, 1087-1100.

Lei, L., Chen, W., Wu, B., Chen, C., & Liu, W. (2021). A building energy consumption prediction model based on rough set theory and deep learning algorithms. Energy and Buildings, 240, 110886.

Leuven, K. (2015). OpenIDEAS. In https://github.com/open-ideas

Levihn, F. (2017). CHP and heat pumps to balance renewable power production: Lessons from the district heating network in Stockholm. Energy, 137, 670-678.

Li, X., & Wen, J. (2014). Review of building energy modeling for control and operation. Renewable and Sustainable Energy Reviews, 37, 517-537.

Li, Z., Han, Y., & Xu, P. (2014). Methods for benchmarking building energy consumption against its past or intended performance: An overview. Applied Energy, 124, 325-334.

Library, S. L. Scikit Learn Library. Retrieved 1 July,2023 from https://scikit-learn.org/stable/modules/svm.html

Lipinski, T., Ahmad, D., Serey, N., & Jouhara, H. (2020). Review of ventilation strategies to reduce the risk of disease transmission in high occupancy buildings. International Journal of Thermofluids, 7, 100045.

Liu, J., Wang, S., Wei, N., Chen, X., Xie, H., & Wang, J. (2021). Natural gas consumption forecasting: A discussion on forecasting history and future challenges. Journal of Natural Gas Science and Engineering, 90, 103930.

London, M. o. (2024). London Heat Map. https://apps.london.gov.uk/heatmap/

Ltd, D. S. Design Builder. Retrieved 1 Sep,2023 from https://designbuilder.co.uk

Lumbreras, M., Martin-Escudero, K., Diarce, G., Garay-Martinez, R., & Mulero, R. (2021). Unsupervised clustering for pattern recognition of heating energy demand in buildings connected to district-heating network. 2021 6th International Conference on Smart and Sustainable Technologies (SpliTech),

Maalka Tool. Retrieved 1 sep,2023 from https://www.maalka.com/tools

Magoulès, F., & Zhao, H.-X. (2016). Data mining and machine learning in building energy analysis. John Wiley & Sons.

Mallala, B., Khan, P. A., Pattepu, B., & Eega, P. R. (2024). Integrated energy management and load forecasting using machine learning. 2024 2nd International Conference on Sustainable Computing and Smart Systems (ICSCSS),

Marino, D. L., Amarasinghe, K., & Manic, M. (2016). Building energy load forecasting using deep neural networks. IECON 2016-42nd Annual Conference of the IEEE Industrial Electronics Society,

Martin, D. (2017). The impact of building orientation on energy usage: Using simulation software IDA ICE 4.7. 1,University of Gävle, Faculty of Engineering and Sustainable Development, Department of Building, Energy and Environmental Engineering, Energy system, Sweden.

Martínez, S., Eguía, P., Granada, E., Moazami, A., & Hamdy, M. (2020). A performance comparison of multi-objective optimization-based approaches for calibrating white box building energy models. Energy and Buildings, 216, 109942.

Mbiydzenyuy, G., Nowaczyk, S., Knutsson, H., Vanhoudt, D., Brage, J., & Calikus, E. (2021). Opportunities for machine learning in district heating. Applied Sciences, 11(13), 6112.

met, E. Envi met. Retrieved 1 Aug,2023 from https://www.envi-met.com/students/

Migilinskas, D., Balionis, E., Dziugaite-Tumeniene, R., & Siupsinskas, G. (2016). An advanced multi-criteria evaluation model of the rational building energy performance. Journal of Civil Engineering and Management, 22(6), 844-851.

Mocanu, E., Nguyen, P. H., Gibescu, M., & Kling, W. L. (2016). Deep learning for estimating building energy consumption. Sustainable Energy, Grids and Networks, 6, 91-99.

Mohammadusman Doddamani, A. (2023). Energy simulations of apartment buildings using IDA ICE. In.

Morille, B., Lauzet, N., & Musy, M. (2015). SOLENE-microclimate: a tool to evaluate envelopes efficiency on energy consumption at district scale. Energy Procedia, 78, 1165-1170.

Moser, A. G. C., Muschick, D., Gölles, M., Lerch, W., Schranzhofer, H., Nageler, P. J., Mach, T., Tugores, C. R., & Leusbrock, I. (2019). Co-Simulation of an Energy Management System for Future City District Energy Systems. 2019 International Conference on Innovative Applied Energy,

MUDIT SAXENA, P. M., INDERDEEP DHIR. Xerohome. Retrieved 1 Oct,2023 from https://www.xerohome.com/about

Mueller, A. C. (2010). Analyses of building energy system alternatives through transient simulation

Murphy, A. R., & Fung, A. S. (2019). Techno-economic study of an energy sharing network comprised of a data centre and multi-unit residential buildings for cold climate. Energy and Buildings, 186, 261-275.

Nazarenko, E., Varkentin, V., & Polyakova, T. (2019). Features of Application of Machine Learning Methods for Classification of Network Traffic (Features, Advantages, Disadvantages). 2019 International Multi-Conference on Industrial Engineering and Modern Technologies (FarEastCon),

NEO Net Energy Optimizer. Retrieved 1 Sep,2023 from https://www.buildingenergysoftwaretools.com/software/neo-net-energy-optimizer®

NREL. Object-Oriented Controllable High-Resolution Residential Energy Model. In https://www.nrel.gov/grid/ochre.html

Ogliari, E., Eleftheriadis, P., Nespoli, A., Polenghi, M., & Leva, S. (2022). Machine Learning methods for clustering and day-ahead thermal load forecasting of an existing District Heating. 2022 2nd International Conference on Energy Transition in the Mediterranean Area (SyNERGY MED),

Open Dataset by Ministry of Housing, Communities & Local Government (https://opendatacommunities.org/home

Open Energy Modelling Framework. In. (2016). https://github.com/oemof

Oró, E., Taddeo, P., & Salom, J. (2019). Waste heat recovery from urban air cooled data centres to increase energy efficiency of district heating networks. Sustainable Cities and Society, 45, 522-542.

Pan, Y., Zhu, M., Lv, Y., Yang, Y., Liang, Y., Yin, R., Yang, Y., Jia, X., Wang, X., & Zeng, F. (2023a). Building energy simulation and its application for building performance optimization: A review of methods, tools, and case studies. Advances in Applied Energy, 10, 100135.

Pan, Y., Zhu, M., Lv, Y., Yang, Y., Liang, Y., Yin, R., Yang, Y., Jia, X., Wang, X., & Zeng, F. (2023b). Building energy simulation and its application for building performance optimization: A review of methods, tools, and case studies. Advances in Applied Energy, 100135.

Patel, P., Sivaiah, B., & Patel, R. (2022). Approaches for finding optimal number of clusters using k-means and agglomerative hierarchical clustering techniques. 2022 international conference on intelligent controller and computing for smart power (ICICCSP),

Perez, D., Kämpf, J., Wilke, U., Papadopoulo, M., & Robinson, D. NEIGHBOURHOOD OF ZÜRICH CITY. Proceedings of CISBAT 2011 - CleanTech for Sustainable Buildings, 937-940. https://infoscience.epfl.ch/record/174435?ln=en

Perez, D., Kämpf, J. H., Wilke, U., Papadopoulou, M., & Robinson, D. (2011). CITYSIM simulation: the case study of Alt-Wiedikon, a neighbourhood of Zürich City. Proceedings of CISBAT 2011-CleanTech for Sustainable Buildings, 937-940.

Petersen, S. H., C.A. (2013). iDBuild. Retrieved 1 Sep,2023 from http://www.idbuild.dk

Pinto, G. building automation energy data analytics (BAEDA). In http://www.baeda.polito.it/research/decision_support_systems_for_building_energy_management

pjayathissa. Rc Building Simulator. In https://github.com/architecture-building-systems/RC_BuildingSimulator

Poponi, D., Bryant, T., Burnard, K., Cazzola, P., Dulac, J., Pales, A. F., Husar, J., Janoska, P., Masanet, E. R., & Munuera, L. (2016). Energy technology Perspectives 2016: towards sustainable urban energy systems. International Energy Agency.

Pro, H. Homer Pro. Retrieved 1 Sep,2023 from https://www.homerenergy.com/index.html

Prokhorenko, V., Choo, K.-K. R., & Ashman, H. (2016). Context-oriented web application protection model. Applied Mathematics and Computation, 285, 59-78.

Radtke, U. (2022). A brief literature review of structuring district heating data based on measured values. The Ohio Journal of Science, 122(2), 75-84.

Rak, K., Völker, J., Taeger, J., Bahmer, A., Hagen, R., & Albrecht, U.-V. (2019). Medizinische apps in der hno-heilkunde. Laryngo-Rhino-Otologie, 98(S 01), S253-S289.

Raka Adrianto, L., Grandjean, P.-A., & Sawalha, S. (2018). Heat recovery from CO2 refrigeration system in supermarkets to district heating network. 13th IIR Gustav Lorentzen Conference,

Rashad, M., Khordehgah, N., Żabnieńska-Góra, A., Ahmad, L., & Jouhara, H. (2021). The utilisation of useful ambient energy in residential dwellings to improve thermal comfort and reduce energy consumption. International Journal of Thermofluids, 9, 100059.

Rehmann, F., Cudok, F., Schölzel, J., Schreiber, T., Henn, S., & Streblow, R. (2024). District Energy Management Systems: Key Data Points for System Integration and Related Challenges: Lessons Learned from Experts in Germany. Energy Technology, 2400297.

REopt. Retrieved 1 sep,2023 from https://reopt.nrel.gov/tool

ResCheck Web. Retrieved 1 sep,2023 from https://energycode.pnl.gov/REScheckWeb/#/login

Romanchenko, D., Odenberger, M., Göransson, L., & Johnsson, F. (2017). Impact of electricity price fluctuations on the operation of district heating systems: A case study of district heating in Göteborg, Sweden. Applied Energy, 204, 16-30.

Rushirajsinh, Z. (2023). The Elbow Method: Finding the Optimal Number of Clusters

Retrieved May, 23th from https://medium.com/@zalarushirajsinh07/the-elbow-method-finding-the-optimal-number-of-clusters-d297f5aeb189

Rusmardiana, A., Akhirina, T. Y., Yulistyanti, D., & Pauziah, U. (2018). A web-based high school major decision support system in Banten using tsukamoto's fuzzy method. 2018 International Seminar on Intelligent Technology and Its Applications (ISITIA),

RWTH Aachen University,TEASER. In. https://github.com/RWTH-EBC/TEASER

Sadeghian Broujeny, R., Ben Ayed, S., & Matalah, M. (2023). Energy Consumption Forecasting in a University Office by Artificial Intelligence Techniques: An Analysis of the Exogenous Data Effect on the Modeling. Energies, 16(10), 4065.

Salkuti, S. R. (2019). Day-ahead thermal and renewable power generation scheduling considering uncertainty. Renewable Energy, 131, 956-965.

Sardoueinasab, Z., Yin, P., & O'Neal, D. (2018). Energy modeling and analysis of inherent air leakage from parallel fan-powered terminal units using EMS in EnergyPlus. Energy and Buildings, 176, 109-119.

Schito, E., & Lucchi, E. (2023). Advances in the optimization of energy use in buildings. In (Vol. 15, pp. 13541): MDPI.

Schuster, D. C. (2020). Developing a Decision-Making Framework for a District Energy System Manager Purdue University].

Seefoo Jarquin, C. S., Gandelli, A., Grimaccia, F., & Mussetta, M. (2023). Short-Term Probabilistic Load Forecasting in University Buildings by Means of Artificial Neural Networks. Forecasting, 5(2), 390-404.

Sefaira. (2015). Sefaira. Retrieved 1 Sep,2023 from https://www.sketchup.com/products/sefaira

SEMERGY - Energy Efficient Buildings. Retrieved 1 sep,2023 from https://www.buildingenergysoftwaretools.com/software/semergy-energy-efficient-buildings

Senapt. In. (2020). https://www.senapt.co.uk

Shahcheraghian, A., Madani, H., & Ilinca, A. (2024). From white to black box models: A review of simulation tools for building energy management and their application in consulting practices. Energies, 17(2), 376.

Sintong, J. E. (2023). Data Centres as Prosumers: A Techno-Economic Analysis. In.

SkySpark. SkySpark. Retrieved 1 Sep,2023 from https://skyfoundry.com/product

Smart Energy Retrieved 1 Sep,2023 from http://smartenergysoftware.com

Snuggpro. Retrieved 1 Sep,2023 from https://snuggpro.com

Sola, A., Corchero, C., Salom, J., & Sanmarti, M. (2018). Simulation tools to build urban-scale energy models: A review. Energies, 11(12), 3269.

Solemma-LCC. Diva , Rhino. Retrieved 1 Sep,2023 from https://www.solemma.com/diva

Speed simulation platform. Retrieved 1 Sep,2023 from https://speed.perkinswill.com

Stef Boesten, W. I., Stefan C. Dekker, and Herman Eijdems. (2019). 5th generation district heating and cooling systems as a solution for renewable urban thermal energy supply. Retrieved May 8th from https://adgeo.copernicus.org/articles/49/129/2019/

Steuer D, A. J., Magnius R, Arias J, Sawalha S. (2023). Techno-economic analysis of heat export from supermarket refrigeration systems: field measurements analysis of three case studies. 26th IIR International Congress of Refrigeration, Paris.

Su, C., Dalgren, J., & Palm, B. (2021). High-resolution mapping of the clean heat sources for district heating in Stockholm City. Energy Conversion and Management, 235, 113983.

Su, C., Madani, H., & Palm, B. (2019). Building heating solutions in China: A spatial techno-economic and environmental analysis. Energy Conversion and Management, 179, 201-218.

Sun, Y., Haghighat, F., & Fung, B. C. (2020). A review of the-state-of-the-art in data-driven approaches for building energy prediction. Energy and Buildings, 221, 110022.

The Swedish Survey Agency. (2024). Retrieved April 16th from https://www.lantmateriet.se/en

Sze, V., Chen, Y.-H., Yang, T.-J., & Emer, J. S. (2017). Efficient processing of deep neural networks: A tutorial and survey. Proceedings of the IEEE, 105(12), 2295-2329.

TEASER Software Project,Institute for Energy Efficient Buildings and Indoor Climate. Retrieved 1 Oct,2023 from http://rwth-ebc.github.io/TEASER/index.html

Tool, C. Cove Tool. Retrieved 1 sep,2023 from https://www.cove.tools

Testasecca, T., Lazzaro, M., Sarmas, E., & Stamatopoulos, S. (2023). Recent advances on data-driven services for smart energy systems optimization and pro-active management. In 2023 IEEE International Workshop on Metrology for Living Environment (MetroLivEnv) (pp. 1–6). IEEE.

Thinsungnoen, T., Kaoungkub, N., Durongdumronchai, P., Kerdprasop, K., & Kerdprasop, N. (2015). The clustering validity with silhouette and sum of squared errors. In Proceedings of the 3rd International Conference on Industrial Application Engineering 2015 (pp. 280–285). Nakhon Ratchasima Rajabhat University.

A TRaNsient SYstems Simulation Program. University of Wisconsin Madison Retrieved 1 Sep,2023 from https://sel.me.wisc.edu/trnsys/

TRNSYS. TRNSYS Software. Retrieved 1 Sep,2023 from https://www.trnsys.com

Tschätsch, C., Turner, E., Marston, A., Zakhor, A., Lemmond, J., & Baumann, O. (2016). Smart energy analysis calculator–an interactive tool for automating building energy analysis & expediting energy audits. Proceedings of SimBuild, 6(1).

University, E. CitySim https://www.epfl.ch/labs/leso/transfer/software/citysim/

University of Strathclyde, E.-r. University of Strathclyde, ESP-r. Retrieved 1 Oct,2023 from https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&ved=2ahUKEwjZ1ObN6rL4AhW8DkQIHSVSBU0QFnoECA4QAQ&url=https%3A%2F%2Fwww.strath.ac.uk%2Fresearch%2Fenergysystemsresearchunit%2Fapplications%2Fesp-r%2F&usg=AOvVaw2xWOd0BpLQelj6bIsYBaqz

University, U. Building Energy and Water Data. Retrieved April 22nd from https://energy.ubc.ca/energy-and-water-data/skyspark/

University, U. BUILDING MANAGEMENT SYSTEMS (BMS). Retrieved April 22nd from https://energy.ubc.ca/ubcs-utility-infrastructure/building-management-systems-bms/

University, U. SKy Spark. Retrieved April 22nd from https://energy.ubc.ca/projects/skyspark/

University, U. (2024). HOT WATER DISTRICT ENERGY SYSTEM. Retrieved july 12th from https://energy.ubc.ca/ubcs-utility-infrastructure/district-energy-hot-water/

Unternährer, J., Moret, S., Joost, S., & Maréchal, F. (2017). Spatial clustering for district heating integration in urban energy systems: Application to geothermal energy. Applied Energy, 190, 749-763.

Urban, B., & Glicksman, L. (2006). The MIT Design Advisor–A fast, simple tool for energy efficient building design. Proceedings of SimBuild, 2(1), 7. https://www.semanticscholar.org/paper/THE-MIT-DESIGN-ADVISOR-%E2%80%93-A-FAST%2C-SIMPLE-TOOL-FOR-Urban-Glicksman/049ecae3c52fe278d4baa5b9dc5a813d5575855c

US Government Open Data Platform (https://catalog.data.gov/dataset?tags=energy&q=building+energy+

Vázquez-Canteli, J. R., Dey, S., Henze, G., & Nagy, Z. (2020). CityLearn: Standardizing research in multi-agent reinforcement learning for demand response and urban energy management. arXiv preprint arXiv:2012.10504.

Velux. Velux, Daylight Visualizer. Retrieved 1 Sep,2023 from https://www.velux.com/what-we-do/digital-tools/daylight-visualizer

Venujkvenk. (2023). Exploring Time Series Data: Unveiling Trends, Seasonality, and Residuals. Retrieved April 25th from https://medium.com/@venujkvenk/exploring-time-series-data-unveiling-trends-seasonality-and-residuals-5cace823aff1

Vlčková, J., Venkrbec, V., Henková, S., & Chromý, A. (2017). Protection of Workers and Third Parties during the Construction of Linear Structures. IOP Conference Series: Earth and Environmental Science,

Wahlroos, M., Pärssinen, M., Rinne, S., Syri, S., & Manner, J. (2018). Future views on waste heat utilization–Case of data centers in Northern Europe. Renewable and Sustainable Energy Reviews, 82, 1749-1764.

Wang, Z., Wang, Y., Zeng, R., Srinivasan, R. S., & Ahrentzen, S. (2018). Random Forest based hourly building energy prediction. Energy and Buildings, 171, 11-25.

WatchWire. WatchWire. Retrieved 1 Sep,2023 from https://watchwire.ai/?utm_campaign=SourceForge&utm_source=sourceforge&utm_medium=website

wattics. wattics. Retrieved 1 Aug,2023 from https://www.wattics.com

Webb, M., Aye, L., & Green, R. (2018). Simulation of a biomimetic façade using TRNSYS. Applied Energy, 213, 670-694.

Wei, Y., Zhang, X., Shi, Y., Xia, L., Pan, S., Wu, J., Han, M., & Zhao, X. (2018). A review of data-driven approaches for prediction and classification of building energy consumption. Renewable and Sustainable Energy Reviews, 82, 1027-1047.

Werner, M., Geisler-Moroder, D., Junghans, B., Ebert, O., & Feist, W. (2017). DALEC–a novel web tool for integrated day-and artificial light and energy calculation. Journal of Building Performance Simulation, 10(3), 344-363.

Werner, S. (2017). District heating and cooling in Sweden. Energy, 126, 419-429.

Wijaya, C. Y. (2019). Breaking down the agglomerative clustering process. Retrieved April 16th from https://towardsdatascience.com/breaking-down-the-agglomerative-clustering-process-1c367f74c7c2

Wu, W., Deng, Q., Shan, X., Miao, L., Wang, R., & Ren, Z. (2023). Short-Term Forecasting of Daily Electricity of Different Campus Building Clusters Based on a Combined Forecasting Model. Buildings, 13(11), 2721.

www.geeksforgeeks.org. K means Clustering – Introduction. Retrieved April 16th from https://www.geeksforgeeks.org/k-means-clustering-introduction/

Zekić-Šušac, M., Has, A., & Knežević, M. (2021). Predicting energy cost of public buildings by artificial neural networks, CART, and random forest. Neurocomputing, 439, 223-233.

Zhang, G., Ge, Y., Pan, X., Afsharzadeh, M. S., & Ghalandari, M. (2022). Optimization of energy consumption of a green building using PSO-SVM algorithm. Sustainable Energy Technologies and Assessments, 53, 102667.

Zhang, K., Zhang, Y., Liu, J., & Niu, X. (2018). Recent advancements on thermal management and evaluation for data centers. Applied Thermal Engineering, 142, 215-231.

Zhang, X., Xia, J., Jiang, Z., Huang, J., Qin, R., Zhang, Y., Liu, Y., & Jiang, Y. (2008). DeST An integrated building simulation toolkit Part II: Applications. Building Simulation,

Zhang, Y., Johansson, P., & Kalagasidis, A. S. (2022). Assessment of district heating and cooling systems transition with respect to future changes in demand profiles and renewable energy supplies. Energy Conversion and Management, 268, 116038.

Zhao, L. P., R.K. (2022). Integration of Machine Learning in Building Energy Models. J. Build. Perform. Simulation, 45, 12.