UNIVERSIDAD PONTIFICIA COMILLAS
Higher Technical School of Engineering ICAI

MATHEMATICAL ENGINEERING AND ARTIFICIAL INTELLIGENCE

Final Degree Project

Filling the gaps for optimal electrification of space conditioning in buildings

author
Adrián López-Lanchares Echezarreta - 202107370

director
Pablo Dueñas Martínez

Madrid, June 2025

Abstract

This project presents a data-driven framework to optimize the electrification of space conditioning in buildings. It proposes clustering models to classify over 39,000 heat pump entries into representative groups and thermal models to estimate building-specific temperature behavior. The models are designed for integration into the DECARB optimization framework, allowing efficient and interpretable decision-making. While deep learning models showed strong predictive performance, a linear regression model was ultimately selected for its compatibility with optimization tools. The results demonstrate that simplification through clustering and interpretable modeling can effectively support scalable building decarbonization strategies aligned with global climate goals.

I hereby declare, under my own responsibility, that the Project presented with the title

Filling the gaps for optimal electrification of space conditioning in buildings

at the High Technical School of Engineering - ICAI of Universidad Pontificia Comillas in the academic year **2024/25** is my own work, original and unpublished, and has not been previously submitted for any other purpose.

The Project is not a copy of someone else's work, neither totally nor partially, and any information taken from other documents has been properly referenced.

Signed: **Adrián López-Lanchares Echezarreta** Date: ....

Project submission authorized
PROJECT SUPERVISOR

Signed: **Pablo Dueñas Martínez** Date: ....

Contents

1 Introduction

Energy consumption in buildings is responsible for 30% of total US greenhouse gas emissions in 2019 [1]. Emissions from buildings can be direct, from the combustion of fossil fuels on-site, or indirect, from the electricity used in the building. Direct emissions accounted for 12.5% [1] of total US greenhouse gas emissions, released from the combustion of fossil fuels for uses such as space heating, water heating, and cooking. Indirect emissions accounted for 18% of total US greenhouse emissions [1].

The biggest contributor to direct emissions from buildings is space heating, responsible for 7% of total US emissions in 2019 [15]. This amount of greenhouse gas emission, equivalent to 410 million tons of CO2, is greater than the entire national emissions of Spain [3].

To decarbonize these buildings, there is a wide consensus that electrification offers the most cost-effective and scalable route (IEA Net Zero by 2050) [5]. Mainly, the replacement of fossil fuel-powered heating equipment with electric alternatives, such as Air Source Heat Pumps (ASHPs). These heat pumps provide both heating and cooling, and could be used to replace conventional fossil-fueled heating equipment, such as boilers and furnaces powered by natural gas, oil, or propane.

While electrification does not completely eliminate emissions, ASHPs offer a significant reduction in emissions compared to their conventional fossil-fueled counterparts. In addition, emissions from electrification will improve as electricity generation becomes cleaner, proving electrification of space heating for residential buildings to be a powerful tool for the reduction of greenhouse gas emissions in the sector.

The electrification of space heating will play a major role in the decarbonization of buildings [5]. Approximately 1.5 million heat pumps are installed every month globally, but this figure should rise to 10 million monthly installations globally, in order to meet the net-zero goal. The surge of heat pump use by 2050 would cause electricity's share in energy consumption for space heating, water heating and cooking to go up from 18% to more than 40% by 2050. In addition to the new heat pump installations, 85% of buildings globally should comply with zero-carbon-ready building energy codes by 2050, meaning fossil fuel use in the building sector must decline by 96%.

1.1 Context and Motivation

In order to optimize the decarbonization of a building, a lot of variables come into play, such as a building's thermal model, the heat pump model or even user behavior. Because of this, powerful mathematical models are needed in order to model and optimize a building's decarbonization.

An important factor such a model should take into consideration is the wide range of heat pumps available, and their importance of outside temperature in order to work at maximum efficiency. Heat pump efficiency can drop sharply in colder climates when outside temperatures fall below the system's optimal design points. The introduction of Cold Climate Heat Pumps (CCHPs) also makes electrification of space heating suitable for regions with harsh winters. A powerful model capable of accounting for all the possibilities is needed in order to optimize the electrification of buildings.

The process of electrification also calls for thermal models of buildings in order to estimate each building's thermal necessities. The thermal properties of a building's materials (e.g., walls, windows, insulation) impact its ability to retain heat, which creates different thermal profiles for different buildings. Building occupancy, thermal mass, building location and climate are other examples that affect a building's thermal needs.

As shown in this chapter, powerful mathematical models capable of processing all of these variables are needed in order to optimize the decarbonization of buildings and to reach the Net Zero goal by 2050 [5].

1.2 Objectives

The main objective is to capture and optimize the dynamics in the replacement of fossil fuels with heat pumps through the use of data-driven models and artificial intelligence techniques. These dynamics include the ones introduced in the previous chapter, such as heat pump model and thermal conditions of the building.

To achieve this, a series of models are proposed to solve each of the aforementioned dynamics, focusing on machine learning approaches to each of them.

Firstly, a data-driven model is proposed for the classification and clustering of heat pump models, focusing on design temperature to ensure the model is as accurate as possible. This model will enable the optimization of heat pump model selection for different climates and building thermal profiles.

Secondly, a heavily data-focused model to assign thermal necessities and thermal profiles to buildings, given each building's different thermal properties. A different model is proposed for each building, tailored to each building's thermal characteristics.

Finally, the implementation of the proposed models in the DECARB optimization framework, a mixed integer linear program developed by researchers at the MIT Energy Initiative, that minimizes a building's total energy costs. Further details on the DECARB optimization framework are provided in Section 2.

Although a machine learning model capable of understanding user behavior was mentioned as an objective in the first exhibit, this surpassed scope, and was left as a separate project for the future, further explored in Section 6.

1.3 Structure of the Paper

1.3 Structure of the Paper

The remainder of this paper is structured as follows:

Section 2, "State of the Art," presents an overview of existing methodologies and frameworks related to building electrification and optimization. It specifically emphasizes the DECARB optimization framework and identifies current limitations that the present study aims to address.

Section 3, "Methodology," describes in detail the methodological approach adopted for this study. It includes two main components: the Air Source Heat Pump (ASHP) clustering model, aimed at simplifying the selection of representative heat pump models; and the building thermal model, developed to accurately predict indoor temperature dynamics.

Section 4, "Experiments," outlines the series of experiments conducted to evaluate the effectiveness and suitability of the proposed models. It elaborates on the objectives, the experimental setups, and the criteria for assessing model performance.

Section 5, "Results," provides a comprehensive presentation of the outcomes obtained from the experimental phase. It covers detailed analyses of the clustering models and the thermal prediction model, highlighting performance metrics and the implications for practical application.

Finally, Section 6, "Conclusion and Future Work," synthesizes the key findings, emphasizes the contributions of this research to the field of building electrification optimization, and suggests potential directions for future investigation and model enhancement.

2 State of the Art

Before embarking on any research project, it is essential to examine existing frameworks and studies that are relevant to the project's scope. Specifically, in the context of electrifying space conditioning in buildings, the **DECARB.jl** framework emerges as a critical foundation. DECARB.jl utilizes mixed-integer linear programming (MILP) optimization to model energy systems comprehensively, covering technologies such as Combined Heat and Power (CHP), HVAC systems, absorption chillers, water heaters, photovoltaic panels, wind turbines, battery storage, and electric vehicle infrastructure.

The strength of DECARB.jl lies in its detailed integration of thermal and electrical modeling, enabling precise operational simulations and effective investment planning for decarbonization. While existing tools like REopt [9] and EnergyPlus [2] provide significant contributions—REopt in multi-technology optimization and EnergyPlus in detailed building simulations—they do not fully address the integrated operational and investment optimization that DECARB.jl offers.

Nevertheless, certain aspects within DECARB.jl still remain underexplored. Notably, there is a significant potential in enhancing the framework through the integration of advanced, data-driven approaches and machine learning models. These include:

- **Optimized heat pump selection based on climatic conditions and thermal characteristics.** The current framework adds a list of equipment without tailoring representative heat pump models to the specific building-climate characteristics.
- **Detailed building thermal model profiling.** The current setup defines each building by a RC thermal model based on constructive characteristics, which has performed poor under real-world simulation conditions.

The present paper addresses precisely these gaps by developing targeted machine learning models intended to enrich the DECARB framework. These models aim to optimize heat pump selections, refine building-specific thermal modeling, and explore preliminary approaches to incorporating user behavior into energy optimization strategies.

3 Methodology

The work of this project has been heavily focused on data-driven models, making use of machine learning techniques to learn new relationships between the used data. Specifically, clustering models were used to group together different models of ASHPs and find representatives of the hundreds of different heat pump models. Finally, a variety of machine learning and deep learning techniques and models were used in order to estimate a building's thermal model. Each of the models' experiments and results will be addressed in the following sections.

All of the project's programming was made using Python 3.10, and popular libraries such as Pandas [12] for data processing, Numpy [11] for linear algebra calculations, Scikit-Learn [14] for clustering models and PyTorch [13] for all remaining models. All of the project's code can be found in the GitHub repository [7].

3.1 ASHP Clustering Model

3.1.1 Model Design

The first objective of this study has been the creation of a machine learning model capable of clustering different Air-Source Heat Pumps (ASHPs) into a given number of distinct groups, each group having one heat pump as its representative. Due to the differences in ASHP design temperature, multiple models were created, each for a specific design temperature. The design temperatures for which clustering models are proposed are -8 °C, 8 °C, 28 °C and 35 °C, as these are the most popular design temperatures for cold and hot climates, respectively. Under mild-temperature conditions, ASHPs either perform efficiently or are not needed. Additionally, two different models are proposed for each ducting configuration. A model for single-zone ducted buildings, and a separate one for multi-zone ducted buildings, to ensure similar heat pumps are still clustered together when building characteristics change. This approach yields a total of eight distinct clustering models, each corresponding to a specific combination of zoning configuration and design temperature. This ensures that the selected ASHP representatives remain valid and representative across a wide range of operational scenarios.

A K-means model is proposed for the clustering of ASHPs. Other clustering techniques were used to ensure the most optimal clustering was created, as described in Section 4. Ultimately, eight K-Means clustering models were developed—two for each of the design temperatures previously discussed. For every design temperature, one model was trained for single-zone ducted buildings and another for multi-zone ducted buildings, allowing the clustering process to account for differences in HVAC zoning configurations.

The main limitation of using K-means as the clustering method stems from the practical requirements of the problem: representative heat pumps must correspond to real models from the NEEP database [8]. However, K-means computes synthetic cluster centroids that minimize the distance between each data point and its assigned cluster center, without regard for the existence of real heat pump entries at those centroids. To address this, an additional selection step is introduced in the training pipeline: for each cluster, the heat pump that lies closest to the centroid is selected as its representative. This approach ensures that the full set of over 39,000 heat pump entries in the NEEP Database is reduced to a smaller, manageable subset, while guaranteeing that each representative corresponds to an actual product in the database.

3.1.2 Training and Validation

In order to train the clustering models, native train functions of the Scikit-Learn [14] library were used, trained to group ASHPs based on various efficiency and capacity metrics. To validate different clustering models, the average distance from a heat pump to its cluster center was used, to judge a model's clustering capabilities. Visual validation was also conducted by applying Principal Component Analysis (PCA) for dimensionality reduction, enabling the projection of clustering results onto a two-dimensional space. This allowed for a qualitative assessment of the clustering model's performance and the degree of separation between clusters.

3.1.3 Datasets

The clustering model was trained using data on various heat pump models, carrying information like heat pump design temperature, and efficiency and capacity measurements at different external temperatures. We used the NEEP Cold Climate Air Source Heat Pump Specification Database [8], developed by the Northeast Energy Efficiency Partnerships, which contained all of the required data to train the model. The NEEP Heat Pump database contains over 39,000 Air Source Heat Pump entries, and a model was trained over this dataset in order to reduce this amount to the requested number of representatives.

3.1.4 Data Processing

To train the heat pump clustering model, a subset of features from the NEEP Database [8] was selected with the aim of achieving accurate grouping of heat pumps while minimizing the dimensionality of the input data. A total of seven input variables were ultimately selected for the clustering model, variables that capture both the energy efficiency and capacity performance of the heat pump across different operating regimes: the **Seasonal Energy Efficiency Ratio (SEER)**, which measures the average seasonal efficiency and reflects real-world performance; the **Coefficient of Performance (COP)** at minimum, rated, and maximum loads, quantifying the efficiency ratio between thermal energy output and electrical energy input at each operating point; and the **minimum, rated, and maximum capacity** values, representing the thermal output in these regimes and describing the unit's flexibility and operational range. Together, these metrics enable the model to accurately distinguish equipment based on their energy behavior and thermal characteristics, resulting in well-separated clusters. The COP and capacity variables were categorized according to design temperature, resulting in the creation of separate datasets for each design condition provided in the NEEP Database. The design temperatures considered in this study were -8 °C, 8 °C, 28 °C and 35 °C. The reason for this selection of parameters lies in a main physical feature of heat pumps: the output capacity and COP declines with the distance between the outdoor and design temperatures.

3.2 Building Thermal Model

3.2.1 Model Design

In order to model buildings' thermal characteristics and make it compatible with DECARB's current thermal model behavior, a machine learning model capable of predicting a building's internal temperature is proposed. This model uses observable data, such as outside temperature, radiation levels, heating and cooling consumption, and previous internal temperature values. This model behaves as a time series forecasting model, predicting the building's internal temperature hourly timeseries one step at a time.

To achieve this, a Linear Regression model is proposed, to estimate each of the input variables' parameters, and subsequently use these parameters in the DECARB optimization framework. Although many deep-learning techniques were experimented with, a Linear Regression model was finally chosen for the problem. These alternative models will be discussed in Section 4.

3.2.2 Training and Validation

Building thermal models, developed using the PyTorch library [13], were trained and validated using custom-made functions, in order to accommodate for different training parameters, optimizers and loss functions, and monitor model training in real time. The final model used for modeling building thermal characteristics was trained for 100 epochs with a batch size of 32, using an Adam optimizer with learning rate of 10^{-4}, and tuned using a Mean Squared Error loss.

For the evaluation of thermal profile prediction models, the dataset was divided chronologically, with the first 80% of the calendar year data used for training and the remaining 20% reserved for validation. This split resulted in slightly more than two months of unseen data for model assessment. The selected validation period was sufficient to evaluate model performance under a variety of seasonal and operational conditions, providing a reliable basis for comparison across the different modeling approaches.

3.2.3 Datasets

Models created to estimate a building's thermal model were all trained using NREL's ResStock dataset [10]. The ResStock dataset contains information on thousands of buildings, includng thermal information like ducting or insulation, as well as climate information, building heating energy consumption, building temperature, orientation. In total, more than 300 informational variables are stored for each building, in addition to multiple time series including hourly heating energy consumption and building indoor temperature. As will be explored in Section 4, different input variables were experimented with, including building layout variables such as building floor area, wall material or number of windows, and building heating energy consumption.

Finally, weather data from ResStock's AMY2018 dataset [10] was used, specifically outdoor temperature data and direct solar radiation data, both in hourly time series format. All used ResStock datasets were datasets for Massachusetts, where the weather data used corresponded with a given building's county inside the state of Massachusetts.

3.2.4 Data Processing

The majority of the data processing in this paper was done using NREL's ResStock dataset [10]. As a first step, all time series were standardized to an hourly resolution, as the original sampling frequencies varied across datasets. These time series include weather data, heating energy consumption data for each building, and internal building temperature.

Weather data played a critical role across all model architectures and experiments, making its proper processing essential. The analysis focused on two key time series: ambient temperature and direct normal radiation, as these variables were identified as the most influential. Preprocessing steps included unit conversions, resampling the data to an hourly resolution, and handling missing values to ensure data quality and consistency.

In addition to time-dependent data, such as time series, the ResStock dataset [10] also provides a wide range of time-independent variables, including building characteristics, insulation levels, and layout information. While these static features were not incorporated into the final linear regression or ARIMA models, they were extensively explored in the development of deep learning approaches, where they were integrated with time series data to enhance model performance. Specifically, time-independent features were aggregated with temporal inputs for training Multi-Layer Perceptron (MLP) and Long Short-Term Memory (LSTM) models. These models are discussed in detail in Section 4.

To predict a building's internal temperature time series using MLP and LSTM architectures, a sliding window technique was applied. Each input sample consisted of a sequence of 12 time steps (equivalent to 12 hours), corresponding to a single target value used for supervised learning. The input sequences included past values of internal temperature, external temperature, direct solar radiation, and heating energy consumption. Additionally, encoded time-independent building characteristics were appended to each sequence. The full list of features used for this purpose is presented in Appendix A.

Since most of the static features were categorical rather than numerical, an encoding step was necessary. Given the relatively low number of unique categories per feature, one-hot encoding was selected to represent non-numeric variables. After processing and encoding the time-independent features, they were concatenated with each temporal window to form the final input used by the deep learning models.

Finally, the proposed linear regression model only takes temporal variables as input, these being: Previous building internal temperature, current and previous external temperatures, direct solar radiation, and heating energy consumption. This results in 5 parameters capable of modeling a building's thermal profile, by modeling its internal temperature. No bias term is used in the proposed model, as this approach works best for the integration with the DECARB optimization framework.

4 Experiments

4.1 Objectives

With the aim of determining the most effective model for each objective, a series of experiments were carried out by changing model architecture, parameters and data processing approaches. Finding the optimal configuration for each model guarantees that predictions are made optimally and errors are minimized.

The primary objective of all the conducted experiments was to select the optimal model architecture for each problem, these being the clustering of ASHPs and the estimation of a building's thermal model. Selecting the optimal model architecture proved to be crucial, not only to minimize the model's error, but to ensure its complexity was low enough, as a complex deep learning model behaving as a black box would not work inside an optimization framework such as DECARB.

Secondary objectives of experimentation included data processing techniques such as data normalization, unit conversions, and a combination of variable selection and variable importance evaluations, in order to determine relevant ResStock [10] features for thermal profile prediction.

4.2 Experiments on Models

4.2.1 Clustering Model

The primary subject of experimentation in this section was the selection of the most optimal clustering model architecture. Two different architectures were selected to study, K-means clustering and DBSCAN clustering, as other options proved to be too complex for the problem at hand.

Both models achieve the final result of clustering a dataset, but operate on fundamentally different principles. K-means clustering relies on one hyperparameter, where the number of desired clusters is indicated. The algorithm starts by assigning random centroids for each cluster, and iteratively updates the centroids until the most optimal clustering is achieved for the given number of clusters.

Conversely, DBSCAN clustering relies on two hyperparameters: the neighborhood radius ε and the minimum number of points minPts. The algorithm groups together points that are closely packed, and labels as outliers those that lie in low-density regions. Unlike K-means, it does not require the number of clusters to be specified in advance and can identify clusters of arbitrary shape.

Experiments using both model architectures and different hyperparameters proved K-means to be the most optimal architecture for the heat pump clustering problem. Selecting the number of clusters showed to be crucial, as keeping this number relatively low was vital to achieve a satisfactory result. The main limitation of the DBSCAN algorithm lies in the need to define a distance-based radius parameter, which can be problematic in high-dimensional spaces, where distance metrics often become less informative and may lead to suboptimal clustering performance.

Regarding parameter selection for K-Means clustering, elbow plots were used to determine the optimal number of clusters by balancing model inertia and the simplicity of the clustering structure.

Finally, Principal Component Analysis (PCA) was explored as a dimensionality reduction technique with the aim of decreasing the complexity of the input data and potentially enhancing clustering performance. However, the results showed no significant improvement, and thus PCA was excluded from the final pipeline in order to maintain model simplicity.

4.2.2 Building Thermal Model

In line with the experiments carried out for the clustering model, model architecture was the primary subject of study for the estimation of a building's thermal profile. Various approaches were used to determine the most optimal model architecture for the problem at hand, ranging from time series forecasting models such as ARIMA, to deep learning models, namely Multi-Layer Perceptron models and LSTM [4] models, optimized for time series prediction. However, due to the complexity limitations the DECARB optimization framework posed, a linear regression model was ultimately selected, in order to estimate a building's thermal profile with a low number of parameters. Additionally, all experiments were carried out for a single building, but the same approach can be followed for all buildings in the ResStock Dataset [10].

The thermal behavior of a building can be modeled using a simplified lumped-capacitance energy balance, which assumes the building's air volume behaves as a single thermal node. It is important to note that indoor temperature is not purely a passive thermal response to external conditions and internal heat inputs. In residential buildings, temperature is typically regulated by a thermostat, which activates the heating system when the temperature drops below a certain threshold and deactivates it when the upper threshold is reached. This results in a control mechanism that produces a step-like evolution of the indoor temperature, rather than a smooth continuous curve. Understanding this behavior is key when modeling or predicting indoor temperature, as it introduces nonlinear dynamics and discontinuities into the thermal profile. The continuous-time equation describing the evolution of the indoor temperature T_{in} is:

$$C \cdot \frac{dT_{in}}{dt} = \frac{T_{out} - T_{in}}{R} + Q_{hvac} + \eta \cdot Q_{solar}$$

Here, C is the building's effective thermal capacitance (in J/K), and R is the thermal resistance between the indoor and outdoor environments (in K/W). The term Q_{hvac} represents the heating or cooling power delivered to the building (in W), while Q_{solar} corresponds to the incident solar radiation (in W), modulated by a gain factor η that accounts for the building's solar heat gain characteristics.

This equation reflects the principle of conservation of energy: the rate of change of internal energy is determined by the net heat exchange through the building envelope, the mechanical heating/cooling provided, and the absorbed solar energy. For practical implementation in a data-driven context, this equation can be discretized and reformulated into a linear regression model to estimate the indoor temperature over time.

$$\begin{aligned} T_{in,t} = \beta_0 &+ \beta_1 T_{in,t-1} + \beta_2 T_{out,t} \\ &+ \beta_3 T_{out,t-1} + \beta_4 Q_{solar,t} \\ &+ \beta_5 Q_{hvac,t} + \epsilon_t \end{aligned} \tag{1}$$

where:

- $T_{in,t-1}$: previous indoor temperature
- $T_{out,t}$: current outdoor temperature
- $T_{out,t-1}$: previous outdoor temperature

- $Q_{solar,t}$: current direct solar radiation
- $Q_{hvac,t}$: current heating/cooling energy consumption
- $\beta_0, \ldots, \beta_5$: regression coefficients
- ϵ_t: error term

Several modeling approaches were explored to estimate a building's indoor temperature, with the goal of identifying the most representative model for capturing its thermal dynamics. The selected models ranged from statistical time series methods to deep learning architectures, each evaluated for accuracy, complexity, and compatibility with the DECARB optimization framework. While complex models offered strong predictive performance, they posed integration challenges due to their black-box nature. Ultimately, a linear regression model was selected for its simplicity and interpretability, but its performance was benchmarked against the more sophisticated alternatives to validate its effectiveness.

The first model evaluated was the Autoregressive Integrated Moving Average with Exogenous variables (ARIMAX) model [6]. This time series forecasting model extends the traditional ARIMA framework by incorporating external variables believed to influence the target time series. In this case, the exogenous inputs included hourly heating energy consumption and outdoor temperature, both introduced with a one-time-step lag to account for temporal dependencies. Direct solar radiation was also included. Despite tuning input selection, lag structure, and model order, ARIMAX failed to adequately capture the building's thermal behavior, prompting a shift toward more flexible modeling strategies.

Deep learning models were subsequently evaluated, starting with the Multi-Layer Perceptron (MLP). MLPs are feedforward neural networks consisting of layers of neurons, where each layer applies a linear transformation followed by a non-linear activation function. Given an input vector $\mathbf{x}$, a hidden layer computes $\mathbf{h} = \sigma(\mathbf{W}\mathbf{x} + \mathbf{b})$, where $\mathbf{W}$ and $\mathbf{b}$ are learnable parameters, and σ is an activation function such as ReLU. Several MLP configurations were tested, varying in depth, activation type, and the use of regularization. The best-performing architecture employed one hidden layer with ReLU activations and no dropout. Figure 1 shows an MLP architecture of one input layer, two hidden layers and one output layer.

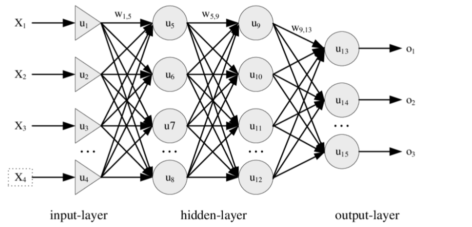

Figure 1: Multi-Layer Perceptron (MLP) model architecture.

In parallel, Long Short-Term Memory (LSTM) networks [4] were explored due to their ability to capture temporal dependencies in sequential data. LSTMs use memory cells and gating mechanisms to manage information flow across time steps. Figure 2 shows the architecture of a single LSTM layer. The final LSTM architecture consisted of two stacked layers with a hidden size of 32 and no dropout. Both MLP and LSTM models received time series inputs that included lagged outdoor temperature and heating energy consumption, along with solar radiation and static building characteristics extracted from the ResStock dataset [10], as detailed in Section 3.2.4. This setup reflects the building's thermal inertia, where past conditions influence current indoor temperature.

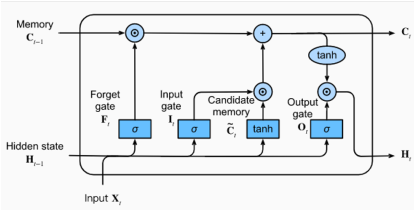

Figure 2: Long Short-Term Memory (LSTM) model architecture.

To support model selection, exploratory feature analysis was conducted to determine the most relevant ResStock variables, with the final selection summarized in Appendix A. While both deep learning models demonstrated promising performance, they were ultimately rejected for use in DECARB due to their black-box nature, which conflicts with the framework's requirement for model interpretability and analytical tractability.

As a result, experimentation focused on linear regression models. These offer full transparency, are easy to implement within an optimization context, and allow for analytical insight into variable influence. The final linear regression model used five input features: current and lagged outdoor temperature, direct normal solar radiation, heating energy consumption, and the previous indoor temperature. This feature set balanced simplicity and predictive power, enabling effective indoor temperature forecasting while satisfying DECARB's constraints. The results, further discussed in Section 5, were comparable to those of the more complex models, demonstrating that linear regression is a viable alternative despite its simplicity.

5 Results

5.1 Heat Pump Clustering Results

As mentioned in Section 3, the chosen model architecture to perform clustering over the Air Source Heat Pump (ASHP) database was K-Means clustering. A total of 8 models are proposed, one for each design temperature (-8 °C, 8 °C, 28 °C and 35 °C), and for each ducting configuration (Single-zone ducting and Multi-zone ducting).

The performance of the clustering models was quantitatively assessed using two key metrics: the average distance to cluster center and the maximum distance to cluster center. These metrics provide insight into the compactness and representativeness of each cluster, with lower values indicating more coherent and tighter groupings of Air Source Heat Pumps (ASHPs).

The following tables summarize the results of the K-Means clustering procedure across all experimental configurations. Table 1 presents the clustering performance for the single-zone configurations, while Table 2 displays the corresponding results for the multi-zone case. Each row within the tables corresponds to a specific design temperature, allowing for a clear comparison of clustering behavior across temperature settings within each zoning category.

Design Temperature (°C)	Avg. Distance to Center	Max Distance to Center
−8	2.64	25.95
8	2.19	18.54
28	1.54	8.40
35	2.25	79.65

Table 1: K-Means clustering metrics for single-zone configuration.

Design Temperature (°C)	Avg. Distance to Center	Max Distance to Center
−8	2.67	12.11
8	2.93	30.94
28	1.18	6.10
35	2.60	51.31

Table 2: K-Means clustering metrics for multi-zone configuration.

The clustering results show that all models achieved relatively low average distances to their respective cluster centers, indicating that the K-Means algorithm successfully grouped heat pump models into compact clusters for each configuration. However, the maximum distances in some cases are significantly higher than the average, suggesting the presence of outliers or instances less well represented by their assigned clusters. This is an expected behavior in clustering large and heterogeneous datasets, and it reinforces the importance of selecting representative heat pumps carefully, particularly those furthest from the centroid. Overall, the results validate the use of K-Means for this application, while also highlighting the need for post-clustering review of extreme cases.

These results show the selected heat pumps are good representatives of the whole dataset, and divide it into optimally-placed clusters. The optimal number of clusters was determined to be five for all models. This choice was guided by an analysis of the elbow criterion, whereby the within-cluster sum of squares exhibits a pronounced inflection at k = 5 in each case. Accordingly, all models were specified to partition the data into five clusters. Figure 3 presents the elbow plot for a specific model, clearly demonstrating the marked “elbow” at k = 5 and thereby validating this configuration.

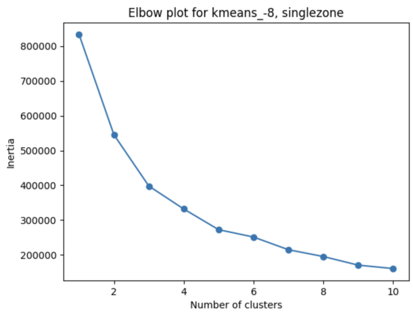

Figure 3: Elbow plot of the K-Means model trained for -8 °C and single-zone ducting configuration.

The representative heat pumps identified through clustering serve as a reduced set for integration into the optimization framework. This approach significantly simplifies the problem, as it allows the original dataset of over 39,000 distinct heat pump models to be condensed into only five representative Air-Source Heat Pump (ASHP) models, thereby reducing complexity without compromising diversity of performance characteristics. These representatives, as well as the whole dataset's clustering results, can be found in Figure 4, referring to a 2D PCA projection of the K-Means clustering result for the −8 °C single-zone configuration.

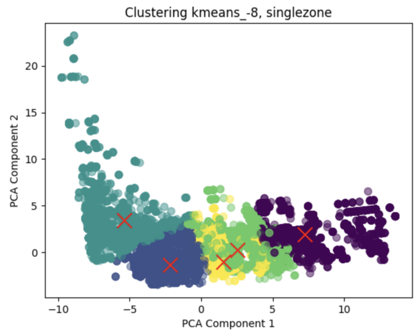

Figure 4: Each point represents a heat pump model, colored by assigned cluster, while red crosses denote the PCA-transformed cluster centroids.

To provide further insight into the technical characteristics of these selected representatives, Tables 3 and 4 present their respective heating capacities and coefficients of performance (COP) at three different operating points (minimum, rated, and maximum), all measured at −8,°C. Table 3 summarizes the thermal output ranges in Btu/h, highlighting the diversity in system sizes captured by the clustering process. Table 4, on the other hand, presents the efficiency metrics of each representative under varying capacity conditions. Together, these tables help illustrate how the clustering retained a broad spectrum of operational characteristics across the original dataset, ensuring representativeness in both size and performance.

SEER	Min. Capacity (Btu/h)	Rated Capacity (Btu/h)	Max. Capacity (Btu/h)
108	3,200	8,800	12,700
20,725	7,600	15,100	15,100
21,773	10,200	23,200	23,200
26,180	14,000	25,200	25,200
39,260	21,860	34,400	34,800

Table 3: Minimum, rated, and maximum heating capacities of the selected heat pump representatives at −8°C and single-zone ducting configuration.

SEER	COP at Min. Capacity	COP at Rated Capacity	COP at Max. Capacity
108	2.61	2.93	2.68
20,725	3.01	2.65	2.65
21,773	2.25	2.56	2.56
26,180	2.63	2.56	2.56
39,260	2.49	2.20	2.21

Table 4: Coefficient of Performance (COP) at minimum, rated, and maximum capacity of the selected representatives at −8°C

5.2 Building Thermal Model Results

A diverse set of models was evaluated for the task of predicting a building's thermal profile, as detailed in Section 4. These include the Autoregressive Integrated Moving Average with Exogenous variables (ARIMAX) model [6], as well as two deep learning architectures: the Multi-Layer Perceptron (MLP) and the Long Short-Term Memory (LSTM) networks [4], along with a baseline Linear Regression model. To estimate the building's internal temperature, each model was trained using a selected set of input features, as outlined in Section 3.2.4, including weather conditions, energy consumption, and prior temperature values.

Although the ARIMAX model demonstrated acceptable predictive performance from a statistical perspective, it fell short in terms of capturing the fundamental thermal behavior of the building. Its forecasts exhibited a rigid and repetitive structure, showing limited responsiveness to variations in input features such as solar radiation or outdoor temperature. This pattern suggests that the model's predictive mechanism primarily relied on short-term autoregressive memory, rather than effectively incorporating the external influencing factors relevant to thermal dynamics. While this approach may result in relatively low error metrics, particularly in datasets with smooth temporal continuity, it does not yield a model that is informative or explanatory. Given that the objective of this work is not merely short-term prediction but the extraction of parameters that characterize the building's thermal inertia and responsiveness, ARIMAX is ultimately unsuitable for the intended purpose.

In contrast, the deep learning models MLP and LSTM demonstrated a substantially better capacity to capture the nonlinear and time-dependent interactions present in the thermal data. Both models were able to learn from sequences of past conditions and internal states, yielding forecasts that more accurately reflected the building's response to environmental and operational variables. The LSTM model, in particular, showed a clear advantage in modeling temporal dependencies, especially over longer time horizons, due to its gated memory architecture. These qualities translated into strong empirical performance, as evidenced by the Mean Squared Error (MSE) values obtained during training and validation.

Model	Mean Squared Error (MSE)
MLP	0.1211
LSTM	0.0446
Linear Regression	0.1472

Table 5: Comparison of MSE values for thermal profile prediction models.

Table 5 summarizes the MSE values for all models under consideration. The results indicate that the LSTM model achieved the lowest prediction error, with an MSE of 0.0446, followed by the MLP at 0.1211. Although the Linear Regression model recorded the highest MSE at 0.1472, this value remains within a reasonable range considering the model's simplicity and the small amplitude of the target variable. The indoor temperature values for the evaluated building ranged approximately between 21°C and 23°C throughout the period studied. Within this narrow range, an MSE of 0.1472 corresponds to a relatively modest deviation, reinforcing the idea that even simple models can provide a meaningful approximation of indoor thermal behavior under certain conditions. Figures 5, 6 and 7 show MLP, LSTM and Linear Regression model predictions, respectively. These results illustrate each model's performance over the test dataset, which, in the case of input data spanning a full calendar year, corresponds to slightly more than two months of unseen observations.

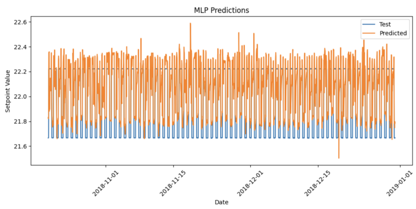

Figure 5: Predictions over the test dataset for the Multi-Layer Perceptron (MLP) model

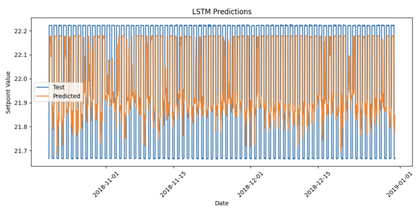

Figure 6: Predictions over the test dataset for the Long Short-Term Memory (LSTM) model

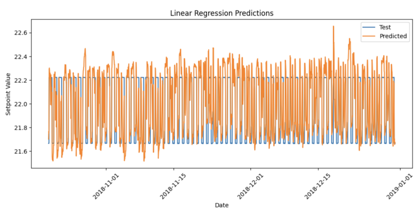

Figure 7: Predictions over the test dataset for the Linear Regression model

Nevertheless, model selection in this study is not driven solely by predictive accuracy. The final choice of Linear Regression is justified primarily by its high level of interpretability and its seamless integration into optimization-based frameworks. Unlike deep learning models, which behave as black boxes and require substantial computational resources, the linear model offers transparency and computational simplicity. These features are particularly advantageous when the model is used within a mixed-integer linear programming environment such as DECARB, where the explicit mathematical formulation of thermal dynamics is a prerequisite. While LSTM and MLP models could theoretically yield better predictions, their complexity and opacity make them impractical for optimization tasks that demand explainable and computationally efficient formulations.

In summary, while all models tested were capable of approximating the building's thermal behavior to varying degrees, the Linear Regression model represents the most balanced choice for the objectives of this work. It combines acceptable predictive accuracy with interpretability, low complexity, and compatibility with downstream optimization applications.

6 Conclusion and Future Work

This study has developed a comprehensive framework to support the electrification of space conditioning in buildings through the use of data-driven methods. The approach is structured around two core contributions: a clustering-based selection of representative heat pump models, and a predictive modeling strategy for estimating building thermal behavior. Together, these components address key limitations in current optimization tools by reducing model complexity while preserving essential technical and thermal characteristics.

The clustering methodology enabled the reduction of a large set of over 39,000 commercially available air-source heat pump models into a limited number of representative units. This simplification proved critical for enabling computationally efficient integration into the DECARB optimization environment. Despite the dimensionality and diversity of the original dataset, the clustering approach maintained sufficient variability to ensure realistic and applicable results in downstream tasks.

In parallel, the study evaluated multiple data-driven approaches for modeling indoor temperature dynamics based on building and weather data. While deep learning architectures such as LSTM and MLP offered superior predictive performance, they lacked interpretability and compatibility with the structure of the optimization problem. In contrast, the Linear Regression model, though less complex, offered adequate accuracy and aligned well with the need for transparency and analytical tractability within the DECARB framework.

Overall, the proposed methodology demonstrates that significant gains in practicality and computational efficiency can be achieved through careful model selection and data preprocessing, without substantial losses in performance. By enabling the use of simplified yet robust representations of both equipment and thermal response, this work contributes to the broader goal of accelerating the decarbonization of building systems through scalable, optimization-ready tools.

6.1 Future Work

The development of a data-driven machine learning model capable of predicting user behavior has emerged as a new line of investigation. User behavior is a critical yet often neglected factor in residential energy consumption, and its inclusion can significantly enhance the realism and accuracy of predictive models. Specifically, the proposed model aims to capture thermostat usage patterns in order to enhance the understanding of occupant-driven thermal demand. Incorporating such behavioral modeling would offer valuable insights for improving the effectiveness of space conditioning electrification strategies.

The development of a user behavior prediction model remains an important avenue for future research, with the potential to significantly strengthen the current optimization framework by accounting for real-world usage variability and demand-side dynamics.

In addition to the proposed user-behavior model, a promising line of future work involves the development of a deep learning model designed to scale the methodology to a large set of buildings. Specifically, the approach will be applied to a dataset containing over 300,000 residential buildings, incorporating the most relevant construction features extracted from the ResStock database [10]. These construction features could include, but not be limited to, the features presented in Appendix A. The goal is to experiment with different neural network architectures, such as Multilayer Perceptrons (MLPs), Long Short-Term Memory (LSTM) networks, or other suitable models, to infer thermal models for buildings that are not explicitly characterized in the dataset, either due to missing construction typology or undefined thermostat setpoints. This model would enable the generalization of the thermal profile estimation process to unseen or partially defined buildings, enhancing the scalability and adaptability of the overall DECARB framework.

References

[1] Kathryne Cleary and Karen Palmer. "Federal Climate Policy 106: The Buildings Sector". In: (2021). URL: https://www.rff.org/publications/explainers/federal-climate-policy-106-the-buildings-sector/.

[2] U.S. Department of Energy. *EnergyPlus*. 2001. URL: https://energyplus.net.

[3] Johannes Gütschow et al. "Country-resolved combined emission and socio-economic pathways based on the Representative Concentration Pathway (RCP) and Shared Socio-Economic Pathway (SSP) scenarios". In: (2021). URL: https://essd.copernicus.org/articles/13/1005/2021/essd-13-1005-2021-discussion.html.

[4] Sepp Hochreiter and Jürgen Schmidhuber. "Long Short-Term Memory". In: (1997). URL: https://doi.org/10.1162/neco.1997.9.8.1735.

[5] IEA. "Net Zero by 2050. A Roadmap for the Global Energy Sector". In: (2021). URL: https://www.iea.org/reports/net-zero-by-2050.

[6] Chaleampong Kongcharoen and Tapanee Kruangpradit. "Autoregressive Integrated Moving Average with Explanatory Variable (ARIMAX) Model for Thailand Export". In: *Proceedings of the International Conference on Applied Economics (ICOAE)*. Elsevier, 2013. URL: https://www.researchgate.net/publication/255731345.

[7] Adrián López-Lanchares. *Space Conditioning Electrification Repository*. 2025. URL: https://github.com/adrianlopezlanchares/Space-Conditioning-Electrification.

[8] Northeast Energy Efficiency Partnerships. *NEEP Cold Climate Air Source Heat Pump Specification Database*. 2023. URL: https://neep.org/heating-electrification/ccashp-specification-product-list.

[9] NREL. *REopt Web Tool*. 2022. URL: https://reopt.nrel.gov/tool.

[10] Elaina Present et al. *ResStock Dataset 2024.1 Documentation*. 2024. URL: https://www.nrel.gov/docs/fy24osti/88109.pdf.

[11] The Numpy development team. *Numpy: Python Linear Algebra Library*. URL: https://numpy.org/doc/stable/.

[12] The Pandas development team. *Pandas: Python Data Analysis Library*. URL: https://pandas.pydata.org/docs/.

[13] The Pytorch development team. *PyTorch: Python Deep Learning Library*. URL: https://docs.pytorch.org/docs/stable/index.html.

[14] The Scikit-learn development team. *Scikit-Learn: Python Machine Learning Library*. URL: https://scikit-learn.org/1.7/index.html.

[15] Parth Vaishnav and Adilla Mulia Fatimah. "The Environmental Consequences of Electrifying Space Heating". In: (2020). URL: https://pubs.acs.org/doi/10.1021/acs.est.0c02705.

A ResStock Dataset Column Names

- `in.bedrooms`
- `in.duct_leakage_and_insulation`
- `in.duct_location`
- `in.geometry_floor_area`
- `in.geometry_stories`
- `in.geometry_wall_type`
- `in.ground_thermal_conductivity`
- `in.hvac_has_ducts`
- `in.insulation_ceiling`
- `in.insulation_floor`
- `in.insulation_foundation_wall`
- `in.insulation_roof`
- `in.insulation_wall`
- `in.occupants`
- `in.orientation`
- `in.roof_material`
- `in.sqft`
- `in.windows`
- `in.window_areas`
- `in.vintage`