EXAMINING THE PERFORMANCE CHANGE OF INVERSE SURROGATE MODELS WITH BUILDING ENERGY MODEL TIME SERIES DATA

Liam Jowett-Lockwood
B.Eng, University of Victoria, 2021

A Thesis Submitted in Partial Fulfilment of the Requirements for the Degree of

MASTER OF APPLIED SCIENCE

In the Department of Civil Engineering

We acknowledge and respect the Lək̓ʷəŋən (Songhees and Esquimalt) Peoples on whose territory the university stands, and the Lək̓ʷəŋən and W̱SÁNEĆ Peoples whose historical relationships with the land continue to this day.

Examining the Performance Change of Inverse Surrogate Models with Building Energy Model Time Series Data

By

Liam Jowett-Lockwood
B.Eng, University of Victoria, 2021

Supervisory Committee

Dr. Ralph Evins, Supervisor
(Department of Civil Engineering)

Dr. Chris Kennedy, Departmental Member
(Department of Civil Engineering)

Abstract

A building Surrogate Model (SM) is a Machine Learning (ML) model trained to reproduce the outputs of a building energy model at a much smaller computational cost. While a SM will traditionally accept Building Energy Modelling (BEM) parameters for its inputs to predict BEM outputs, a building Inverse Surrogate Model (ISM) suggests doing the opposite. Inverse modelling provides potential in determining unknown building thermal characteristics of existing structures. The task of deriving inputs from outputs is more difficult as multiple input combinations can result in the same output, thereby necessitating the need for comprehensive outputs allowing for more information to be extracted. With the rise of deep learning models and methods, ML practitioners have a greater array of tools available to handle increasingly complex tasks. This has enabled ISMs with a stronger opportunity to excel in parameter prediction.

The papers in this thesis focus on the ability of the ISMs to accurately predict parameter values. The first paper (Chaper 2) examines prediction performance of an ISM with synthetic data from a BEM model based on a single-family home. Performance changes were investigated when data was decreased by reducing the amount of time series provided, the duration of time series, or both. The second paper (Chapter 3) primarily focused on the generalizability of ISMs to be applied for multiple projects without having to retrain on new data each time. Several different ISM models were tested with predicting parameters for different BEM building shapes with varied geometry in addition to multiple locations.

The key finding of this research is that there is potential for ISMs to be used with building data. While all data used in this thesis was synthetic data generated from BEM simulation runs, ISMs were shown to not only successfully predict some parameters, but also hold solid degree of generalizability depending on the ML model used. If ISMs can successfully predict characteristics of an actual building, then it allows for new approaches for applications such as retrofit planning.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

ANN……...Artificial Neural Network

BEM……..…….Building Energy Model

CV (RMSE)……...Coefficient of Variation of Root Mean Square Error

CNN……...Convolutional Neural Network

GHG…… Green House Gas

GRU…… ….Gated Recurrent Unit

HVAC…….Heating Ventilation and Air Conditioning

ISM……..Inverse Surrogate Model

LSTM……....Long-Short Term Memory

ML…….….Machine Learning

MAE……..…..…Mean Absolute Error

MAPE…….Mean Absolute Percentage Error

MSE…….Mean Squared Error

R^2……...Coefficient of Determination

RNN……...…Recurrent Neural Network

RMSE……..….Root Mean Square Error

RMSPE……...Root Mean Square Percentage Error

SM…….....Surrogate Model

Author Contributions

This thesis is comprised of two submitted journal papers (Chapter 2 & Chapter 3) as well as two conference papers (Appendix B & C). The conference papers present earlier versions of the research in Chapter 2. They were peer reviewed and presented at eSim 2022 in Ottawa Canada and Building Simulation 2023 in Shanghai, China. The paper presented at eSim 2022 won the award for "Best Student Paper".

Individual contributions by the authors were the same across each study. Liam Jowett-Lockwood contributed to methodology, software development, writing and editing. Ralph Evins provided supervision, resources, validation and review and editing of the written work. Citations of each paper are:

Chapter 2: Jowett-Lockwood, L & Evins, R. *Using an Inverse Surrogate Model for Determining Building Thermal Characteristics of a Single-Family Home Energy Model.* Submitted to the Journal of Building Performance Simulation.

Chapter 3: Jowett-Lockwood, L & Evins, R. *Examining the Generalizability of Inverse Surrogate Models for Geometrically Non-uniform Building Models.* Submitted to Energy and Buildings

Appendix B: Jowett-Lockwood, L & Evins, R. (2022) *Using a Convolutional Neural Network to Determine the Thermal Characteristics of a Building.* 12th IBPSA eSim Building Simulation Conference

Appendix C: Jowett-Lockwood, L & Evins, R. (2023) *Inverse Surrogate Modelling to Determine Thermal Characteristics of Buildings.* 18th IBPSA International Building Simulation Conference and Exhibition. Shanghai, China.

Acknowledgements

As I look back on the years it took to complete this journey, I realize that it is simply impossible for this list to be small.

I would first like to thank Dr. Ralph Evins for accepting me as a graduate student and providing me invaluable supervision, resources and support throughout. Along with Ralph, I would like to thank everyone at the Energy in Cities group. While lately I haven't been able to spend as much time with you all as I wished I could have, I still cherish many of the memories I received when I lived in Victoria. I would additionally like to thank both IESVic administrative assistant Pauline Shepard and the Civil Engineering graduate secretary Lorrie Barth for their assistance with helping me deal with all the paperwork as well as being extremely patient with me returning signed forms much later than I should have.

I would further like to thank Terry Bergen and Dr. Mohammad Fakoor for providing me an amazing co-op work term at Read Jones Christoffersen Ltd. The industry knowledge I gained proved to be incredibly insightful.

Back at home, I would like to thank both family and friends for their immense support. I couldn't have done it without you all.

Lastly, I would last like to thank Dr. Paul Westermann and Matthias Welzel for their work on convolutional neural networks in the building energy domain. Their research formed the backbone from which my research was allowed to flourish.

Chapter 1 Introduction
1.1 Background

As buildings produce a substantial amount of (Green House Gas) GHG emissions, a focus must not only be given to new constructions, but existing buildings as well. While tearing down an energy inefficient building and replacing it with one that is more efficient remains an option, building energy retrofits are a strong alternative. Building energy retrofits involve improving various characteristics of the building, such as increasing the level of insulation or making it more airtight. Aside from being less expensive, retrofits require less time and materials than building a completely new structure in addition to a smaller embodied carbon footprint.

Deciding on how to implement the retrofit is a necessary step to ensure that it is effective. This requires developing a thorough understanding of the building itself and what aspects of it are the most in need of improvement. In reality, information of a building's thermal properties that help govern its energy consumption, internal temperatures, etc. may have been lost or adjusted overtime through deterioration (e.g. increased air leakage).

A building energy model is a computerized model that provides a representation of the energy performance of a new or existing structure. Inputs into the building energy model can include building envelope properties and equipment loads, while simulated outputs can consist of the energy demand or internal temperature fluctuations overtime. In practice, a building energy model can be calibrated to match the outputs of an existing building which can possibly provide insight into the exact nature of its properties.

While promising, calibrating a building energy model can be a lengthy iterative process involving subsequent simulation runs in order to find a respectable fit. Lately, building surrogate models have gained attention as a means to reproduce the outputs of a comprehensive building energy model at a fraction of the computation cost. These surrogate models are machine learning models that, when trained, forgo the extensive physics computations and instead make accurate predictions. Building inverse surrogate models are the inverse of a traditional building surrogate model whereas they instead provide near-instantaneous accurate predictions on the input data

when provided the output data of a building energy simulation. Successful development of inverse surrogate models would provide practitioners a new non-destructive method of determining of building characteristics, for purposes such as building energy retrofits, fault detection as well as overall owner reassurance.

1.2 Overview of the ISM Process

As a inverse surrogate model in our context is a machine learning model designed to determine BEM inputs from lengthy outputs, its development process can be structured into three sections: data generation, data analysis and model development. The structure is demonstrated in Figure 1.1.

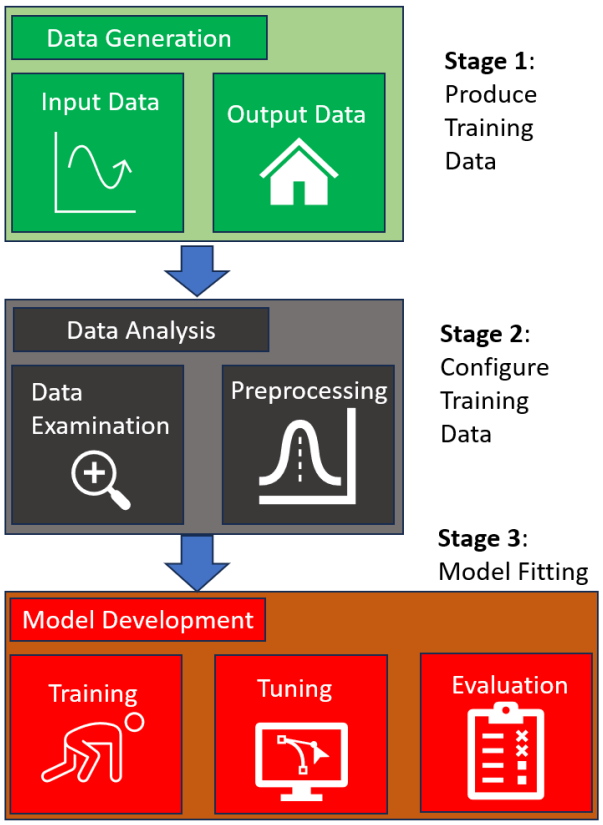


Figure 1.1 Inverse surrogate model creation process. *Developing an ISM can be broken down into three phases, which include acquiring the necessary training data, processing the data and then training, turning and evaluating the model.*

Data generation: As inverse surrogate models are machine learning models, they require a sufficient amount of data to train properly. For real buildings this presents a problem as acquiring enough comprehensive data could not only be costly, but also time consuming. Fortunately,

building energy modelling provides a solution by allowing for vast quantities of data to be produced synthetically. It is important that that appropriate input and output data be selected at this stage as one needs to be sufficiently influential to the other for the inverse surrogate models to effectively make predictions. Output data in the form of various time-series is a suitable input for inverse surrogate models. Building energy model parameters represent influential characteristics of the building in energy performance and therefore represent the ideal prediction for inverse surrogate models.

Data analysis: Prior to training, the data should be examined with appropriate adjustments made. Time-series data can be adjusted in number, length and quality. Quality in this regard pertains to specifically selecting portions of the data for which the inverse surrogate models will benefit the most from training. Observations should also be made for any missing or erroneous data and methods should be applied for handling them. Furthermore, data preprocessing in the form of standardization should be considered to allow for the largest training potential.

Model Development: An appropriate machine learning model should be chosen for the inverse surrogate model so that predicting parameters can be made as accurately as possible. In the same light, the model construction and its hyperparameters also require careful consideration. If possible, hyperparameter tuning can be used to aid this decision making, however the process can often be lengthy. Upon completion of model training, performance evaluation is required, which is often completed with the use of statistical error metrics.

1.3 Outline of the Thesis

Chapter 2 examines the effectiveness of an inverse surrogate model in predicting parameter values when trained on internal temperature time series data from a building energy model based on a single-family home. It examines the changes in prediction performance when the number of time series provided is reduced, the length of the all the time series is decreased as well as when only instances of temperature decay are provided. The inverse surrogate model was composed of a convolutional neural network to best handle the extensive time series data fed into it. By basing the building energy model on a single-family home, the intention is to examine the practicality of parameter prediction with an inverse surrogate model when eventually applied to

real collected data from a similar building. Results indicated that the convolutional neural network was appropriate for accurately predicting building envelope parameters. Depending on the parameter, data limitations could range from highly to minimally impactful. For future work, if the study can be extended to incorporate real data, then it would provide practitioners an alternate means of determining the characteristics of a building.

Chapter 3 focuses on the generalizability of inverse surrogate modelling. While building surrogate models produce outputs substantially faster than their building energy model counterparts, the need to produce vast quantities of training data can greatly offset this benefit. Having an inverse surrogate model be sufficiently generalizable would allow it to be applied to multiple projects without the need to regenerate new data. Several machine learning models were examined which consisted of a simple deep artificial neural network, a recurrent neural network, a convolutional neural network and a transformer. They were all trained on synthetic data from various building energy models which each had their geometry scaled. Additionally examined was training with varied locations as well as on data from all building energy models at once. As with chapter 1, results showed that the inverse surrogate models were successful at accurately predicting parameter values depending on the parameter. This is extended to cases of varied locations as well as combining the data from all building models, though expected performance decreases were observed. Overall, this chapter demonstrates that inverse surrogate modelling is sufficiently generalizable depending on the machine learning model, predicted parameter and scenario.

The papers comprising Chapters 2 and 3 are regarded as separate and are each intended to be submitted for publication in an academic journal. For this reason, there is a degree of overlap between the two papers in regards to introductory matter and some aspects of methodology.

Chapter 4 provides conclusions for the entire work, while also suggesting avenues for future work. Additionally, provided in this thesis are two conference papers detailing earlier work contained in Chapter 2. They are provided in Appendix A and Appendix B respectively. It should be noted that as the work in the conference papers was earlier work, the results have since been improved upon (in Chapter 2) through means such as updating both the machine learning and building energy models.

Chapter 2 Using an Inverse Surrogate Model for Determining Building Thermal Characteristics of a Single-Family Home Energy Model

2.1 Abstract

For countries around the word to achieve their emissions targets, buildings have become a significant focus for energy efficiency. As many homes are now aging, building energy retrofits are a suitable means of lowering operational energy while not drastically increasing embodied carbon. To ensure that retrofits are correctly targeted, it is essential that the properties of the existing building are assessed.

This study examined the use of an Inverse Surrogate model to identify building properties from time series of temperatures based on a residual one-dimensional Convolutional Neural Network. Synthetic training data was collected from a building energy model of an existing building with 14 thermal zones. The model was trained on multizone temperature time-series data with durations of day, week or month. Additionally, the model was trained on instances in time composed solely of temperature decay periods. For each of the different quantities of data, four additional scenarios were examined that involved either randomizing the time interval of data collection, reducing the number of zones and/or in the case of training only on periods of temperature decay, reducing the time-series duration.

Results suggest that prediction performance varied greatly depending on the parameter and time duration. Some parameter predictions were more robust than others to changes in the data quality. Prediction performance typically decreased more when switching to a randomized time interval than lowering the number of zones. Prediction performance noticeably improved when switching from day to week durations, however there were diminishing returns when switching from week to month. When trained on only decay curve instances, prediction performance was strong and experienced more resilience compared to the other tests involving randomized time intervals.

2.2 Introduction

2.2.1 Background

Sustainable and energy efficient buildings are necessary to address energy consumption and greenhouse gas emissions worldwide. In developed countries, over 70 percent of their electrical power generated is used by buildings with 40 percent of their carbon dioxide emissions being produced from the combustion of fuel needed [1]. In order to reach the Government of Canada's 2030 enhanced Paris Agreement target, the Canadian Government has examined buildings as one of many focal points for potential savings reductions [2]. In particular, the 2030 Emissions Reduction Plan specifies investments worth millions of Canadian dollars aimed at helping develop programs to fund both new developments and retrofits of existing buildings [3]. New construction is a focus of this, as unbuilt buildings have the most flexibility when it comes to building materials, mechanical systems, envelope considerations (e.g., amount of insulation to use), etc. However, retrofits are especially important given the ageing building stock in many Canadian cities. For example, in the City of Victoria, 63% of homes were constructed prior to 1981 and 30% were constructed prior to 1961 [4]. While building age alone cannot serve as a perfect indicator for building energy performance (as buildings could have been upgraded overtime), potential for long-term deterioration and originally lower standards suggest poor performance.

A simple solution is tearing down the existing building and constructing a more efficient replacement. Alternatively, building energy retrofits can often provide a more efficient and less costly option. It has been shown that building energy retrofits have a significantly reduced embodied carbon [5]. It is important, however, that the building is assessed properly prior to the retrofit being undertaken. If assessed incorrectly, then a retrofit could be completed where one aspect of the building is improved but the energy performance remains poor as a different unchanged aspect of the building was more influential [6]. Furthermore, older existing buildings may not have the necessary information (e.g., wall insulation thermal resistance, infiltration flow rate) readily available, as it may have been lost over time or altered from deterioration. Determining building properties to assist in retrofit planning should be seen as an essential step in order to prevent inappropriate retrofit measures from being applied.

2.2.2 Data-driven Methods of Identifying Building Properties

This study focuses on only data-driven methods of determining building properties. Conventional non-data-driven methods exist, but they can often be costly, intrusive to occupants, time consuming and labour intensive [7]. For example, an Air-Tightness Test to determine infiltration requires professionals to access the building, the setup of equipment, the completion of the test and the removal of equipment. While reliable, the effort, cost and time required can be prohibitive.

While data-driven methods are not perfect, they can often include added flexibility. Data-driven methods involve a period of data gathering from the existing building that can include energy use data or temperature readings from select rooms. Afterwards a model is fitted to these data readings to produce an estimate of various properties.

Section 2.2.3 covers the decay curve method which demonstrates a simple though limited analytical approach to determining building properties. Section 2.2.4 discusses the concept of inverse modelling which uses Building Energy Modelling (BEM) outputs to derive inputs that frequently consist of building properties. Section 2.2.5 briefly describes the inverse modelling approach of calibrating a BEM model. Section 2.2.6 explains the potential of introducing Machine Learning (ML) to inverse modelling to help offset some of difficulties of other inverse modelling approaches.

2.2.3 Decay Curve Method

We first review the decay curve method, one of the simpler data-driven methods of determining envelope properties by fitting thermal decay curves. This method uses the fact that when the heating system turns off and the outdoor temperature is lower than the indoor temperature, the temperature drop takes the form of an exponential decay curve. Equation 2.1 gives the thermal energy balance [8]; T_{in} is the internal temperature, $\dot{Q}_{in}$ is the heat flow from internal gains, $\dot{Q}_h$ is the heat provided by the heating system, $\dot{Q}_{sol}$ represents the heat provided via solar gains, $\dot{Q}_{ven}$ is the heat flow as a result of ventilation. The thermal resistance of the building is the variable R, which represents the ability to resist changes in temperature via conduction. The variable C represents the thermal capacitance of building in its ability to store energy.

$$C\frac{dT_{in}}{dt}(t) = \dot{Q}_{in}(t) + \dot{Q}_h(t) + \dot{Q}_{sol}(t) - \frac{1}{R}\left(T_{in}(t) - T_{ext}(t)\right) - \dot{Q}_{ven}(t) \quad (2.1)$$

$$C\frac{dT_{in}}{dt}(t) = \dot{Q}_{other}(t) + \dot{Q}_h(t) - \frac{1}{R}\left(T_{in}(t) - T_{ext}(t)\right) \quad (2.2)$$

$$RC\frac{dT_{in}}{dt}(t) = R(\dot{Q}_h(t) + \dot{Q}_{other}(t)) - \left(T_{in}(t) - T_{ext}(t)\right) \quad (2.3)$$

Equation 2.2 groups all heat flow variables, aside from $\dot{Q}_h(t)$, into one variable defined as $\dot{Q}_{other}(t)$. If an assumption can be made that $\dot{Q}_{other}(t)$ is sufficiently small compared to $\dot{Q}_h(t)$, then it can be ignored and equation 3 can be further simplified. Ideally the variable $\dot{Q}_{other}(t)$ can be determined through knowledge of the internal gains considering only the nighttime period to ignore solar gains. As noted in [8], with no heating or additional heat flows and a constant outdoor air temperature, then equation 4 is an analytical solution to equation 2.3.

$$(T_{in}(t) - T_{ext}) = T_{in}(0)\, e^{\frac{-t}{RC}} \quad (2.4)$$

Equation 2.4 represents an exponential decay curve that a building may experience, likely at night when the heating system is turned off and there are low internal gains and zero solar gains. Using a fitting method, it is possible to extract RC, however, it is not possible to separate either R or C individually. This is a substantial hinderance in determining R as the accuracy will depend on an assumption of the value of C. We are primarily interested in determining R in order to understand the current level of insulation in the building and hence the potential benefits of retrofit methods.

Fitting equation 2.4 can be a simple means to acquire information regarding building characteristics with only temperature time series data and relevant assumptions. We later apply this concept together with more advanced inverse modelling approaches.

2.2.4 Inverse Modelling

With a steady increase of the energy consumption in buildings over the last few decades, BEM has grown as a method to achieve energy efficiency through comparing different designs, systems and subsystems, demonstrating acceptance with energy standards and deriving energy

budgets [9]. BEM involves creating a detailed computer model of the building where building properties are inputted as variables referred to as parameters and extensive physics computations are completed to determine various outputs that can include energy consumption and time specific internal temperatures. BEM can be applied to existing buildings as well but must be calibrated if not all parameters are known.

Unlike forward modelling where parameters are provided to a predeveloped system (such as with traditional BEM) as inputs to produce outputs, for inverse modelling, the system is developed with unknown parameter values and known outputs where the parameters must be adjusted to find the best match [10]. Inverse modelling can consist of a variety of methods including the approach of fitting equation 2.4 to determine RC. Whereas the forward approach would be using RC to form the decay curve, the inverse approach uses an already obtained decay curve to determine RC.

2.2.5 Building Energy Model Calibration

This involves the use of on-site output data (electrical consumption, interior temperature, etc.) collected from the actual building and the BEM input parameters are adjusted until a sufficient match of the output data is obtained. Often the purpose behind model calibration is to help rectify the issue of BEM having potential significant differences between output of a building energy model and its actual building and therefore increase the model credibility [11]. BEM parameters include various characteristics or properties of the building, such as the wall insulation conductivity or infiltration flow rate.

Like the decay curve method, the calibration approach can present notable flaws. Namely, it is heavily reliant on the building energy model itself. While it is often assumed that a parameter combination for which the building energy model outputs match the actual building are close to their actual parameter values, this may not be true if there are other factors influencing the model. Additionally, If the number of readings from the actual building is insufficient, then the number of potential parameter combinations could increase implying that a larger number of incorrect combinations may produce results that seem plausible. Furthermore, not only does building the energy model require significant time to construct, but running many simulations required to match the outputs can be prohibitively time consuming.

Model calibration is a common form of inverse modelling, however its difficulties make the adoption of inverse modelling a challenge. While other methods of inverse modelling exist, as shown in section 2.1.7, methods of inverse modelling that do not rely on extensive calibration are often specialized (i.e., are focused on finding a very small number of parameters), such as the decay curve method. Alternatively, the introduction of ML has the potential to provide newer inverse modelling methods the ability to better solve the time consumption problem as well as allow for determining a greater number of parameters.

2.2.6 Inverse Surrogate Modelling

A building energy surrogate model is a ML model designed to replicate the outputs produced from a traditional building energy model when provided inputs in the form of BEM parameters [12]. The Surrogate Model (SM) has the advantage that (once trained) its predicted outputs are produced practically instantaneously. This helps overcome the time required for traditional building energy models, however acquiring training data and then training the SM can still be a lengthy process.

While a SM accepts BEM inputs as inputs and attempts to predict the same outputs, the idea of an Inverse Surrogate Model (ISM) is to reverse this process, as for inverse modelling. An ISM would therefore take in BEM outputs as input to predict the original BEM inputs. We apply this here using temperature time series as inputs to the ISM and corresponding building parameters as the outputs. It should be denoted that this cannot be achieved directly with BEM, as building properties cannot be calculated from temperature data, but instead requiring a time-consuming iterative calibration process.

2.2.7 Related Studies

This section, reviews several approaches to inverse modelling for a variety of purposes, including improvements to BEM calibration and new inverse modelling methods to determine parameters directly. Some approaches examine methods incorporating SM, while others use more conventional methods based on equation 2.4.

It is common for studies to focus on improvements for BEM calibration by applying inverse modelling to assist in determining building parameters. Nagpal, et al. examined introducing surrogate modelling as a solution to the high computational time of iterative BEM

runs during the calibration process [13]. They first trained either a Random Forest or an Artificial Neural Network (ANN) for which an optimization function would be implemented to find the most optimal set of parameters to complete the calibration process. The authors conclude that their technique could be successfully applied to provide reasonable estimates of parameter values. Herbinger et al. utilizes a surrogate ANN for calibration purposes [14]. What makes their study unique is that the surrogate itself calibrates unknown building parameters through gradient descent, whereas traditionally an external optimization algorithm is used instead with the machine learning model. They compared their ANN to another surrogate based powerful black box optimizer and found that, in most instances, their ANN outperformed the other model.

Improved retrofit planning has additionally been explored with the use of inverse modelling. To determine appropriate retrofit measures, Ascione et al. developed two ANNs to be used together [15]. The first ANN was used to predict the primary energy consumption for space heating and cooling as well as the percentage of annual discomfort hours. Afterwards, upon implementation of several energy reduction measures that could be put in place as a part of a retrofit, the second ANN predicts the same output parameters to provide a comparison between the non-retrofitted building and its retrofitted version. Their strategy aims to reduce the computational cost of non-machine learning approaches in determining retrofit effectiveness. Sharif and Hammad introduced a surrogate ANN to aid in determination of optimal retrofit measures in terms of Total Energy Consumption, Life Cycle Cost and Life Cycle Assessment [16]. Their approach trains an ANN on a dataset of retrofit scenarios from a multi-objective optimization model. They take advantage of the improved computation time of an ANN while not significantly sacrificing optimization capability. Asadi et al. applied an ANN to multi-objective optimization for the purposes of retrofit planning [17]. Their findings showed that using an ANN to speed up the optimization process demonstrated suitable prediction accuracy, however the acquisition of training data was time consuming. Shadram et al. explored a multi-objective optimization approach for determining the benefit of operational energy savings as a result of retrofit measures against the increase of embodied carbon [18]. Their findings suggested that the energy savings of performing a retrofit are greater than the increase of embodied carbon.

Aside from improved BEM calibration and retrofit planning, other studies have examined different inverse modelling approaches of parameter identification. Rasooli and Itard explored

deriving global thermo-physical characteristics (split into 4 parameters) via a simple inverse RC model using air temperature and heating consumption data [10]. Their findings suggested that data granularity and season of collection were influential in determination of the values. Gurney et al. examined the ability of three change point models in their ability to predict heating loads and infiltration [7]. The three models differ in data availability; if more temperature trend data is available, different models can be used. The results of the model were mostly consistent across all three and sufficiently similar to results from a physics-based model. Ko and Park utilized transfer learning to better predict the nominal cooling coefficient of performance for electric heat pumps, wall U-value and lighting power density for existing buildings [19]. They first begin training an ANN on synthetic energy simulation data and then transfered the training process to a set of 49 existing buildings with results further evaluated on 12 remaining buildings. Hong and Hoon Lee developed an inverse modelling approach to determine zone air infiltration rate and internal thermal mass by inversing the zone air heat balance equation with zone air temperature [20]. The intention of their methods was to improve building energy model simulation results by providing more accurate information of the building to help match existing buildings.

2.2.8 Areas of Improvement

Aside from improving the calibration process, many examples in literature focus on determining one or more parameters based on specific models from which parameter values can be extracted. This study examines the potential of an ISM in determining multiple parameters when provided BEM output data. Taking inspiration from the decay curve method, as input to the ISM, we provide temperature time series data collected in 10 minute intervals from various rooms. The intention is that the decay curves and other temperature fluctuations will be learnable by the ISM given their relation to building properties. Our method is unique in that we utilize a deep learning Convolutional Neural Network (CNN) for our ISM in order to handle the sometimes extensive time series data. Furthermore, this study examines the ability of the robustness of the ISM to make accurate predictions with varying quantities and adjustments of trainable data. This includes both the duration (length of time series), number of locations in the building from which time series data is acquired and the exact interval of data capture (e.g., instead of always having data from the same month, the month may vary). This will help identify the necessary quantity of information required for suitable performance of the ISM.

This study is a continuation of previous research presented at the eSim 2022 [21] and Building Simulation 2023 [22] conferences. The main additions in this study are an analysis of prediction performance with varying the time interval of data collection, training models only on instances of temperature decay, and predicting multiple parameters at once.

2.3 Methodology

When developing a surrogate model for a task, the process can be broken up into three general sections:

1. Generation of training data and data organization.
2. Construction of the ML model and its training.
3. Evaluation of model performance.

The first of these steps focuses not only on the training data itself and pre-processing it, but additionally firmly establishing the problem in terms of what inputs will provided to the model for it to predict selected outputs. In our case multi-zone temperature time series data is used to predict various building parameters. We examine model zoning and time duration to explore the effect it has on model predicted accuracy. The second section involves the determination of the CNN's layers and decisions regarding hyperparameters. These can be highly influential in the ability of the model to train properly. For example, a poor selection of hyperparameters could lead to underfitting of the model where the model is unable to achieve its maximum potential performance. Alternatively, the model could experience overfitting where the model learns non-intentional information that exist by coincidence in the training set but is not contained in the testing set. This can be seen indicated by a notable discrepancy between prediction performance on the training set and test set.

The third section uses statistical error metrics to provide performance results. Aside from being able to help compare the model used in this study to those in other studies, multiple different error metrics can help provide insight into different aspects of model performance (e.g. higher errors in some metrics more than others may indicate the presence of occasional outliers).

As different parameters affect the internal temperature of the building more greatly than others, it can be expected that the ability of the ISM to make accurate predictions on them will vary. Furthermore, the quality and quantity of input data may also have varying influence. As shown in Figure 2.1, this study examines three different time intervals comprised of daily, weekly and monthly time series data, as well as time intervals composed only of decay curves (time intervals of constant temperature decrease). The intention of adjusting the length of the timeseries is to determine how much data is necessary to make accurate predictions depending on the parameter. Our initial examination holds the time interval consistent, so that the ISM is able to precisely train on data from the exact same day, week or month. Further examination involves randomizing the time interval to provide insight into the ability of the ISM to determine parameters during different parts of the year. Lastly, it can be assumed that the portions of data that are most valuable for training are when the building temperature is either increasing or decreasing, as the rate of temperature change will be influenced by the parameter values. For this reason, training on only decay curves will help provide insight for both the necessity of having non-decay times in the data and the physical dataset size (i.e., number of data points) for accurate prediction performance.

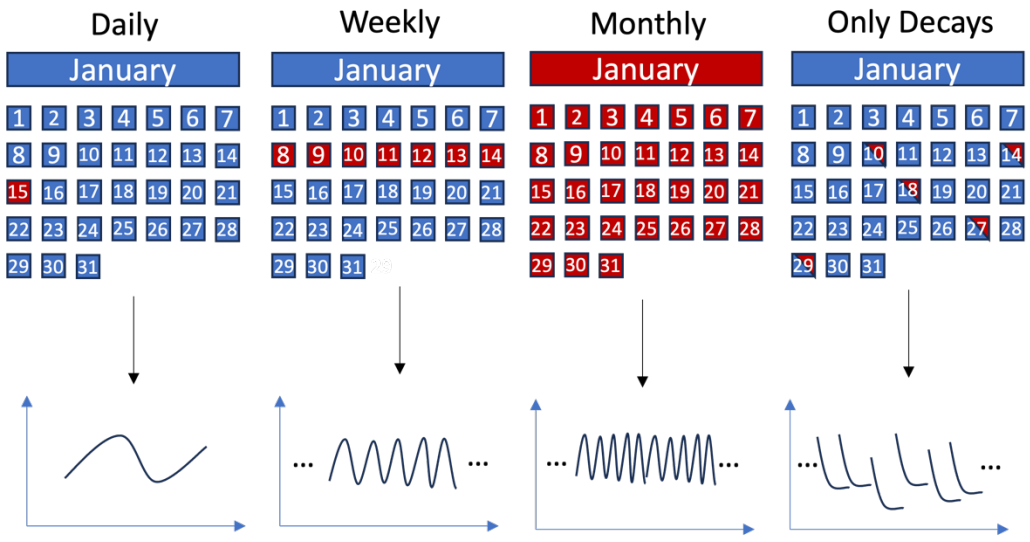
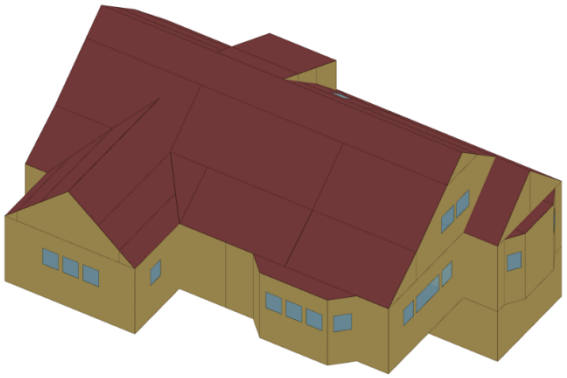


Figure 2.1 Data quantity variability. *The dates shaded in red represent the specific date that data is taken from when the time interval is consistent. Decay curves are taken on dates that experience a period of consistent temperature decrease which can form a suitable curve.*

2.3.1 Building Energy Model

As the building energy model is used to generate training data, it is essential that the model is an accurate depiction of the type of building that we want to predict parameters for. In this study, our building energy model is representative of an aging single-family home located in Victoria, Canada for which local weather data was used. While this study only focuses on test results on synthetic data provided by the building energy model, later the research will include real data measured in an occupied dwelling.

The geometry of the model is shown in Figure 2.2; simulations computations were completed using EnergyPlus [23]. A total of 14 thermal zones were modeled, which each output their own temperature time series. This is broken down into 7 zones on the first floor, 5 zones on the second floor as well as a single zone for each of the basement and attic areas. A heating setpoint schedule was defined as initially 18°C until 06:00, then an increase to 20°C until 09:00, then a decrease back to 18°C until 18:00 and then 20°C until 24:00. Other schedules were held constant for simplicity, such as lighting and equipment energy usage.



Floor	Thermal Zones
Level 3	Attic
Level 2	Master Bedroom Upstairs Bathroom Additional Bedroom Landing Area Walk-in Closet
Main Floor	Family Room Kitchen Dining Room Living Room Entry Area Downstairs Bathroom Bedroom
Basement Level	Basement

Figure 2.2 Building energy model geometry and zone breakdown. *The 14 zone building energy model used in this study is shown on the left with the breakdown of each zone shown in the table on the right.*

No specific Heating, Ventilation and Air Conditioning (HVAC) system was modeled for the home. Instead the EnergyPlus Ideal Air Loads System component was used to represent an

individual highly efficient system, to avoid the complexity of an HVAC system model [24]. Default methods of ventilation and infiltration modelling were used. While the flow rates are parameterized (see section 2.3.2), the actual rate is not affected by external factors such as wind. While somewhat unrealistic, maintaining ventilation and infiltration in this simple fashion should benefit prediction performance as their influence on the multi-zone temperatures will not be affected by unseen weather.

2.3.2 Training Data Generation

The generation of training data involves both the production of inputs to be fed into the machine learning model as well as the outputs the model will try to predict. Each combination of outputs and inputs constitutes one sample. For each of the 14 zones present in the building energy model, temperature time series data is outputted in 10-minute intervals constituting 52,560 total values for the entire year for each zone. In addition to the 14 zones, outdoor air temperature was provided when the interval was randomized (i.e., when the exact day, week or month was not held consistent). This was not necessary for when the interval was consistent as the outdoor air would remain the same across all samples, hence it had no relevance on training. Furthermore, binary data for when heating was applied was also given to assist the ISM by learning when zones were being heated and when not.

The predicted parameters are shown in Table 2.1 along with their value ranges. Parameter value ranges were based on those from previous research [25] and engineering judgement. Parameters which are more highly influential on the overall temperature will likely be more easily predictable than those which are not. Therefore, we would expect that parameters that influence the envelope (e.g., infiltration flow rate) to be more meaningful than small internal gains (e.g., lighting energy). Furthermore, parameters that influence individual locations more easily than others (such as the Foundation Floor Thickness) may enable the ISM to more easily identify them, as their influence will be more specifically concentrated. This could prove to be a hindrance when the number of zones is reduced when predictability of these parameters may more easily suffer.

Table 2.1 Variable parameters and their value ranges. *Parameter selection and value ranges were selected based on engineering judgement and previous research.*

Parameter	Unit	Range
Upper Wall Insulation Conductivity	W/mK	0.01-0.1
Lower Wall Insulation Conductivity	W/mk	0.01-0.1
Roof Insulation Conductivity	W/mk	0.01-0.1
Foundation Floor Thickness	m	0.05-0.2
Glass Conductivity	W/mk	0.005-0.02
Attic Conductivity	W/mk	0.01-0.1
Lighting Energy	W/m^2	5-10
Equipment Energy	W/m^2	5-10
People Quantity	$Person/m^2$	0.015-0.05
Ventilation Flow Rate	L/sPerson	5-10
Infiltration Flow Rate	W/m^2s	0.1-1
Maximum Heating Air Flow Rate	L/s	1-10

Instead of parameterizing the thermal resistance (R-value), conductivity (or thickness for the foundation floor) was selected instead. Parameterizing one of these for a given construction can be considered equivalent to parameterizing the R-value as the thermal resistance is composed of both conductivity and thickness. Additionally, the conductivity for the upper and lower wall constructions (based on building levels) was split to see if the ISM would be able to predict one better than the other. The main differences between the upper and lower floor zones are the glazing area and that the upper floor zones are affected by attic and roof insulation conductivity and lower floor zones are affected by the foundation floor thickness.

The Maximum Heating Air Flow rate parameter was included to limit the effectiveness of the Ideal Air Loads System. Without limitations, the high efficiency would cause the model to respond to changes in the heating setpoint over time exceptionally quickly. While the intention of the ML model is to train based on changes in multi-zone temperatures, rapid changes are prohibitive.

A total of 5,000 samples were produced for this study. 1,000 of these samples were randomly selected to remain in the test set. A further 20 percent of the remaining 4,000 training samples were also randomly selected for the validation set, which then left 3,200 samples purely for training. Whereas the training set is used with the purpose of enabling the model to find connections between the inputs and outputs, the validation set is used during the training process to help prevent overfitting. A validation set functions similarly to the test set in this regard, however the validation set is used during training unlike the test set which is used only once training has completely finished [26].

Latin Hypercube Sampling (LHS) was used to generate sample values. LHS is a widely used sampling method in part because of its ability to significantly reduce unwanted variance for a multitude of applications [27]. By applying LHS, the intention is to ensure that the design space is captured in the data in a uniform manner, so that there is a diminished probability of the ISM being biased by non-uniformly distributed random samples.

2.3.3 Data Organization

Once the training data was acquired it was organized for each of the scenarios outlined. Each sample contained 10-minute interval temperature time series data for each zone for an entire year, however only a fraction of the data (either daily, weekly, monthly or just decay intervals) was used for training. Training on an entire year of data would result in a substantially longer training time and may prove to be unnecessary if it can be demonstrated that accurate predictions can be made on smaller quantities of data.

When the time interval was not randomized, the month of January, its second week and the 15th day were selected for training on month, week or days periods (Figure 2.1). January was selected as it can be anticipated that colder periods will benefit training ability as there will be larger potential for temperature decay (resulting in more notable differences for the ISM to pick up on) when compared to warmer seasons. When the time of year was randomized this was not applied and instead duration was calculated for days and weeks as the number of 10-minute time periods for a given duration. For a day this was 144 (6x24) and for a week this was 1,008 data points. As different months have a different number of days, the total number of 10-minute time steps in a year (52,560) was divided by 12 resulting in a duration of 4,380 steps. To properly

compare, the duration of 4,380 steps was still used when the monthly time interval was held the same even though the month of January has a larger number of time steps.

For training purely on decay curves, the data organization was completed differently. It was necessary to examine instances which held consistent decay over a time period as these would be expected to be the most influential. Instances where this is not achieved and the temperature drop rapidly reaches the new thermal setpoint would provide more limited data and hinder model training. Starting from the beginning of the year, the data was scanned to find suitably long decay curves to be included in the sample. Each decay curve was captured for a two-hour duration regardless of whether or not the decay reached a lower setpoint value and stabilized. For simplicity only temperature data for the upstairs master suite was scanned. Based on the location of this zone, when this zone experiences long decay curves, other zones will likely experience the same phenomena. For each sample, the total number of 2-hour decay curve data points was 1,008 which was the same as 1 week period of data. This totaled 84 decay curves for each zone. As with previously comparing the ISM performance on different durations of data (monthly, weekly and daily), examining the ISM performance only on decay curves was conducted for both the full 1,008 interval of data and half of it to attempt to understand if a reduced time duration would be noticeably impactful. When trained only on decay curves, outdoor air temperature was again provided. Even though most decay curves would occur at the same instance in the data, as the heating setpoints were not modified, there may still be occasions where a sample has a low heating air flow rate value and therefore may have more suitable decay curves.

Lastly, preprocessing was applied to both the training and testing data prior to ISM training. Preprocessing the data is commonly performed prior to training a ML model as they can suffer training difficulties if the features in the data do not resemble standard normally distributed data [28]. Normalizing data in this fashion involves transforming the features such that they each have a mean of 0 and a variance of 1. As the features in the input in this study are the multi-zone temperature time series data, each individual temperature value for each zone is normalized with its counterparts in other samples. The predicted parameters constituting the output data are also normalized. Once the ISM has been trained and has made predictions, the predictions are unnormalized prior to model evaluation.

2.3.4 Convolutional Neural Network Composition

One of the most important layers in a typical ANN is the Fully Connected or Dense layer. Each neuron in this layer is connected to all the neurons in the previous layer thereby providing substantial training ability due to the shear amount of information processing this provides. With sizeable inputs however, this can create issues, such as high training time and memory usage, which weakens simple ANNs as a suitable model for large input tasks. CNNs help rectify this issue via the use of convolutional kernels [29]. Equation 2.5 demonstrates the method of convolution for the first example in Figure 2.3 where the rest of the process is demonstrated for one- and two-dimensional convolution.

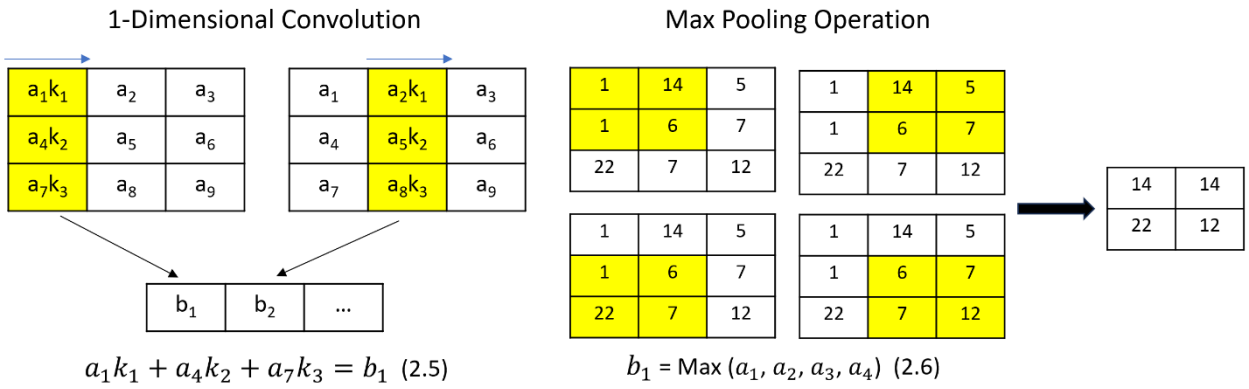


Figure 2.3 One-dimensional Convolution and Max Pooling operations. *For the convolution, the kernel Matrix (yellow highlighted) slides along left to right. For the pooling operation, the pool group (yellow highlight) selects the largest member which is then selected for the output matrix. Variables in equation 2.6 were omitted in the example for clarity.*

While the kernel matrix can be the same size of the input matrix, often the input matrix is larger and therefore requires that the kernel matrix shift along. What is key is that the values in the kernel matrix remain constant across shifting such that the weights are effectively shared. As the kernel weights are used throughout the entire input, the training time can be substantially lower than what would be needed in a Dense layer.

The output matrix is referred to as a Feature Map, which is then typically inputted into a pooling layer to further limit the data size. Different pooling layers exist with Max Pooling being common where the largest value is selected out of a group as shown in Figure 2.3. The use of convolutional and pooling layers helps minimize the data quantity needed, allowing for ease of training while still retaining high prediction accuracy.

Unlike two-dimensional CNNs, which are frequently used for images and video, we opt for a one-dimensional CNN for this study given that our inputs are effectively one-dimensional time-series from multiple zones. One-dimensional CNNs have less computational complexities than their two-dimensional counterparts, therefore allowing their architectures to achieve high performance without being too deep and require less extensive computing hardware to train efficiently [30].

The composition of the CNN used in this study (Figure 2.5) is heavily based on the residual one-dimensional CNN used in [31]. The CNN is composed of three blocks each featuring a convolutional layer followed by a batch normalization layer and then a Rectified Linear Unit (ReLU) activation layer. The latter two layers can be viewed as replacing the pooling layers that would normally be present after a convolutional layer, however an average pooling layer remains near the end of the model. Batch Normalization provides a means of reparametrizing a deep network so that updates between model layers do not conflict with each other, while ReLU is a common activation function where inputs less than zero give zero as the output while positive values remain unchanged [26]. Each block is connected by a residual connection that provides outputs from a previous block to the output of the next block. The intention is to help prevent the vanishing gradient problem from occurring when the depth of the model increases.

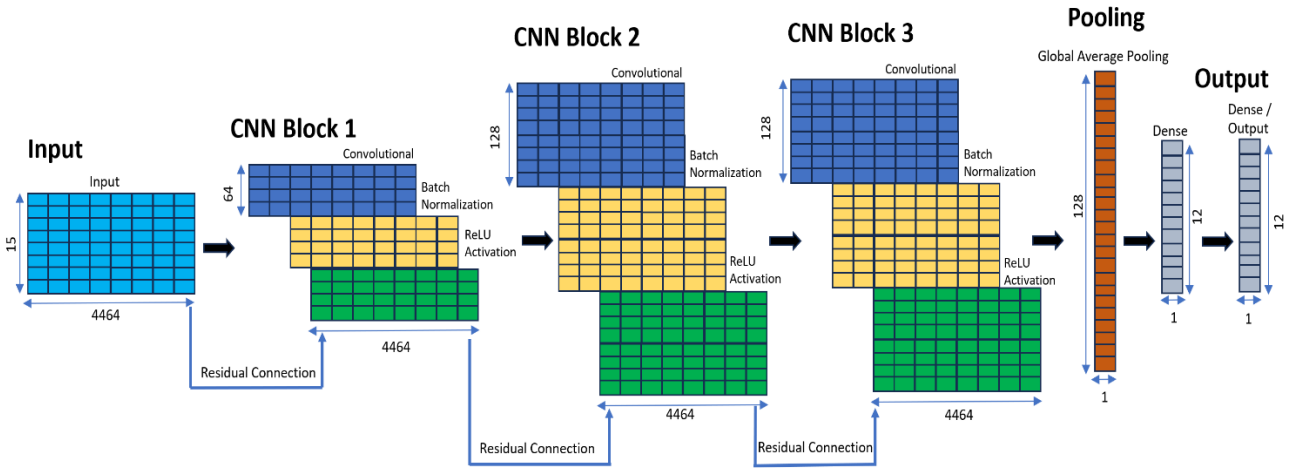


Figure 2.4 One-dimensional residual convolutional neural network. *Each CNN block is composed of three combinations of Convolutional, Batch Normalization and ReLU Activation layers. Residual connections are provided to connect the input from the first layer to the output of the last layer of each block.*

2.3.5 Methods of prediction performance assessment

Assessing ISM performance involved the use of several quantitative error metrics. Error metrics are statistical equations that can be used to effectively assess the ability of a machine learning model in making predictions, assuming a sufficient amount of data is available, which should be true if a portion of the training data is allotted to a test set. Different error metrics should be used to help provide different assessments of model performance. In this study, we use the Coefficient of Determination (R^2), Mean Absolute Percentage Error (MAPE) and the Coefficient of Variation of the Root Mean Square Error (CV[RMSE]), defined as follows:

$$R^2(y, \hat{y}) = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \qquad (2.7)$$

$$MAPE(y, \hat{y}) = 100\left(\frac{1}{n}\sum_{i=1}^{n}\frac{|y_i - \hat{y}_i|}{y_i}\right) \qquad (2.7)$$

$$CV(RMSE)(y, \hat{y}) = \frac{100}{\bar{y}}\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \qquad (2.8)$$

In the above equations, y_i represents the actual values, $\hat{y}_i$ represents the predicted values and $\bar{y}$ is the mean of the actual values. The R^2 error metric indicates how well the predictions effectively fit with the actual data where values typically range from zero to one with values of one indicating a perfect fit and values of zero showing no fit at all. While counterintuitive, it remains possible for it to take on negative values which demonstrate that the model error is higher than the baseline error [32]. Given the typical bounds of the metric, one its most appealing aspects is the ease of comparison across multiple studies [33].

The Root Mean Square Error (RMSE) provides the standard deviation of the differences between the two sets of values [32]. While RMSE is commonly used, arguments have been made against is usefulness as a metric in comparison to others such as Mean Absolute Error (MAE) [34]. Therefore, MAPE was included to provide an additional perspective on performance. One

of the advantages of MAE is its simplicity and ease of understanding and with MAPE being able to be viewed as a percentage version of the MAE, it still retains the ease of understanding while helping it further differentiate itself from RMSE. Furthermore, the quadratic term in RMSE, can suggest the presence of outliers if RMSE is high and MAPE is low [34].

In this study we use the Coefficient of Variation of RMSE (CV [RMSE]) instead of RMSE as CV (RMSE) is a relative metric making it more suitable since the parameters have different units. The CV (RMSE) assess the fit of the model by measuring offsetting error between the two datasets and has been used in several different standards and guidelines such as ASHRAE Guideline 14 [35][36].

2.4 Results

The results are divided into four sections based on the time-duration of data (month, week and day), as well as decay curve durations. In the first three sections, 4 different scenarios were examined: Consistent Time 14 Zones (14-C), Randomized Time 14 Zones (14-R), Consistent Time 5 Zones (5-C) and Randomized Time 5 Zones (5-R). For the reduced zones scenarios, only data from specific zones (main floor bedroom, kitchen, living room, master bedroom and upstairs washroom) were provided to the model. It is anticipated that with removing the consistent day of training and by reducing the number of zones, that overall prediction performance will decrease. For the results with decay curve durations, reducing the number of zones was again considered. This resulted in the 4 scenarios of Full Week 14 Zones (14-F), Half Week 14 Zones (14-H), Full Week Reduced Zones (5-F) and Half Week Reduced Zones (5-H). All 8 total scenarios are further outlined in Table 2.2.

Table 2.2 Result scenarios and descriptions. *The first four relate the month, week and day scenarios while the last are only for the decay curve scenarios.*

Scenario	Description
14-C	• Not applicable for decay curves. • Contains all zones including binary heating data. • Consistent month, week or day.
14-R	• Not applicable for decay curves. • Contains all zones including binary heating data. • The month, week or day is randomized.

5-C	• Not applicable for decay curves. • Contains only 5 zones. • Consistent month, week or day.
5-R	• Not applicable for decay curves. • Contains only 5 zones. • The month, week or day is randomized.
14-F	• Only applicable for decay cures. • Contains all zones including binary heating data. • Contains a full week of decay curves.
14-H	• Only applicable for decay cures. • Contains all zones including binary heating data. • Contains half a week of decay curves.
5-F	• Only applicable for decay cures. • Contains only 5 zones. • Contains a full week of decay curves.
5-H	• Only applicable for decay cures. • Contains only 5 zones. • Contains half a week of decay curves.

2.4.1 Month Results

Error metric scores for each scenario and parameter are shown in Table 2.3.

Table 2.3 Error metric scores for month results. *Performance tends to be high for most parameters with the Infiltration Flow Rate and Maximum Heating Air Flow Rate being the strongest. Prediction performance decreases as the number of zones is reduced and date becomes randomized.*

MONTH	R^2				MAPE				CV (RMSE)			
Scenario	14-C	14-R	5-C	5-R	14-C	14-R	5-C	5-R	14-C	14-R	5-C	5-R
Upper Wall Insulation Conductivity	0.89	0.86	0.61	0.39	18.4	19.7	32.9	36.2	16.2	17.8	29.9	37.5
Lower Wall Insulation Conductivity	0.92	0.95	0.96	0.91	14.7	11.2	11.1	13.9	13.4	10.2	9.5	13.9
Roof Insulation Conductivity	0.96	0.92	0.96	0.79	9.9	13.4	10.0	24.6	9.6	13.4	9.8	22.2
Foundation Floor Thickness	0.95	0.88	0.81	0.73	7.0	10.5	13.9	15.3	8.2	12.3	15.4	18.6
Glass Conductivity	0.91	0.88	0.93	0.74	8.0	11.3	8.7	14.9	10.4	12.0	9.3	17.7
Attic Insulation Conductivity	0.84	0.69	-0.01	-0.07	19.8	26.8	51.9	62.5	18.4	26.0	46.7	48.0
Lighting Energy Power Density	0.37	0.46	0.44	0.31	13.6	12.3	12.7	14.5	15.6	14.5	14.7	16.2
Equipment Energy Power Density	0.50	0.51	0.39	0.23	11.5	11.4	13.1	14.9	13.5	13.4	14.9	16.8
People Quantity	0.79	0.72	0.31	0.16	13.5	14.6	19.3	23.9	14.4	16.5	25.8	28.6
Ventilation Flow Rate	0.88	0.74	0.43	-0.36	5.0	7.6	12.8	20.1	6.5	9.5	14.1	21.9
Infiltration Flow Rate	0.97	0.97	0.99	0.92	12.2	8.5	4.8	11.4	8.5	7.5	5.4	13.2
Maximum Heating Air Flow Rate	0.98	0.92	0.97	0.92	8.9	12.8	8.4	15.4	7.5	13.4	8.2	13.9
Range	0.00		1.00		0.0		70.0		0.0		50.0	

Based on Table 2.3 the higher performing predictions are the Lower Wall Insulation Conductivity, Roof Insulation Conductivity, Infiltration Flow Rate and Heating Air Flow Rate where the R^2 score is above 0.9 for 14-C and decreases are less substantial for other scenarios. The Upper Wall Insulation Conductivity, Foundation Floor Thickness and Glass Conductivity

additionally perform strongly in the 14-C scenario, but have a more significant drop in prediction performance for the other scenarios. For the Upper Wall Insulation Conductivity and Foundation Floor Thickness, as these parameters are more impactful to specific zones (zones in the upper floor and zones impacting the foundation respectively), the prediction performance for them can understandably be more greatly impacted when some of these impactful zones are removed. Both the Infiltration Flow Rate and Heating Air Flow Rate differ in that they are more resilient to changes to the dataset and even during the 5-R scenario, their predictability remains solid.

Predictions of other parameters tend to either have a decent prediction for 14-C and quickly worse for other scenarios, or only moderate to poor prediction performance overall. While decent initially, the Attic Insulation Conductivity, People Quantity and Ventilation Flow Rate parameters notably suffer a significant decrease in their R^2 score when the number of zones is reduced. As the attic is one of the zones removed, the Attic Insulation Conductivity losses its most influential zone and its values are no longer accurately predictable. As the Ventilation Flow Rate is connected to the People Quantity parameter, the ISM may have difficulty discerning them. This is less of an issue when the ISM is provided all zones, as the included basement is underground and receives no natural ventilation, therefore a distinction can be made. Removing this zone partly results in the much weaker prediction performance in the 5-C and 5-R scenarios.

Both the prediction performance on the Lighting Energy Power Density and Equipment Energy Power Density parameters have poor R^2 scores however their performance with the other error metrics remains comparable to the higher performing parameter predictions. This suggests that while the ISM is unable to accurately pinpoint exact values, there remains a decent consistency in prediction accuracy.

2.4.2 Week Results

Error metric results are shown in Table 2.4. Overall prediction performance tends to decrease among each parameter with a few exceptions in the 14-C scenario.

Table 2.4 Error metric scores for week results. *Prediction performance (with a few exceptions) decreases compared to when a month of data is provided. The randomness during training is a possible explanation for the exceptions.*

WEEK	R^2				MAPE				CV (RMSE)			
Scenario	14-C	14-R	5-C	5-R	14-C	14-R	5-C	5-R	14-C	14-R	5-C	5-R
Upper Wall Insulation Conductivity	0.91	0.74	0.81	0.37	18.6	28.4	25.7	44.3	14.5	24.3	21.0	38.0
Lower Wall Insulation Conductivity	0.97	0.90	0.96	0.85	9.3	16.0	9.0	17.6	8.2	15.3	9.6	18.6
Roof Insulation Conductivity	0.96	0.90	0.95	0.81	10.9	15.9	10.9	19.9	9.3	15.4	10.8	21.1
Foundation Floor Thickness	0.84	0.57	0.60	0.24	14.6	21.2	19.7	24.3	14.4	23.3	22.6	31.1
Glass Conductivity	0.96	0.70	0.95	-0.34	6.1	18.4	6.5	43.9	7.1	19.0	7.7	40.4
Attic Insulation Conductivity	0.83	0.49	0.00	-0.17	20.2	33.1	61.1	57.8	19.3	33.1	46.5	50.3
Lighting Energy Power Density	0.44	0.27	0.43	0.29	12.6	14.0	13.0	14.1	14.7	16.8	14.7	16.5
Equipment Energy Power Density	0.50	0.31	0.53	0.23	11.8	13.5	11.3	14.4	13.5	15.9	13.1	16.8
People Quantity	0.82	0.54	0.61	-0.18	12.1	17.8	19.3	35.0	13.3	21.1	19.5	33.7
Ventilation Flow Rate	0.77	0.61	0.68	-0.11	7.8	9.1	7.8	17.5	9.0	11.8	10.5	19.7
Infiltration Flow Rate	0.98	0.97	0.97	0.84	5.2	7.7	7.2	17.0	5.9	7.5	7.8	18.3
Maximum Heating Air Flow Rate	0.98	0.92	0.99	0.88	8.2	14.4	6.5	18.3	6.8	13.9	5.8	16.5
Range	0.00		1.00		0.0		70.0		0.0		50.0	

Compared to Table 2.3, the four highest parameter predictions remain the same with decreases in prediction performance having similar trends to before. Predictions for the Foundation Floor Thickness and Glass Conductivity both suffer significantly lower R^2 scores in the 5-R scenario (though the MAPE score is still decent for the Foundation Floor Thickness) suggesting that they are less resilient when trained on non-consistent and reduced data. For the remaining parameters, prediction performance expectedly decreases in most cases.

It can be observed that performance decreases are more significant with randomizing the week of data collection than reducing the number of zones, with the exception being the Attic Insulation Conductivity parameter. Given that the number of zones is reduced by more than half in the 5-C and 5-R scenarios, this demonstrates that data quality in this case study is notably more important to the ISM than quantity. There are also a few instances where prediction performance actually increases compared to Table 2.3. As these differences are typically small, they could most easily be attributed to variations in model training as the training set is shuffled before each epoch. Lastly, the People Quantity and Ventilation Flow Rate have a slightly improved prediction performance in the 5-C scenario compared to the monthly results and are more accurate compared to the 14-R scenario. Even though the basement zone is omitted, there may exist instances where the ISM is able to discern key characteristics in the data leading to semi-accurate predictions. Prediction performance for these appear unstable given that Table 3 had worse results for the 5-C scenario (likely due to variances in training) and the results for the 5-R scenario were again similarly poor.

2.4.3 Day Results

As with the change from Table 2.3 to Table 2.4, the prediction performance decreases overall, albeit more significantly (Table 2.5). Both the Upper Wall Insulation Conductivity and Glass Conductivity parameters now have completely unsatisfactory performance with all metrics in the 14-R scenario. While high performing for the month and week results, the Lower Wall Insulation Conductivity, Roof Insulation Conductivity and Heating Air Flow Rate parameters have not only a much lower R^2 score, but also significantly higher MAPE and CV (RMSE) errors implying a much greater uncertainty in prediction accuracy. Prediction performance with the 5-C scenario still remains high for these parameters further indicating that randomizing the date on which data is collected is highly impactful.

Table 2.5 Error metric scores for day results. *A substantial decrease is prediction performance is observed when compared to the results with a month and week. This is most noticeable in the 14-R and 5-R scenarios.*

DAY	R^2				MAPE				CV (RMSE)			
Scenario	14-C	14-R	5-C	5-R	14-C	14-R	5-C	5-R	14-C	14-R	5-C	5-R
Upper Wall Insulation Conductivity	0.85	0.03	0.65	0.13	18.2	53.6	32.4	62.2	18.5	47.3	28.5	44.8
Lower Wall Insulation Conductivity	0.92	0.66	0.95	0.58	13.6	30.2	9.8	41.2	13.8	27.7	10.7	31.0
Roof Insulation Conductivity	0.94	0.59	0.93	0.33	13.0	37.2	13.0	38.9	11.8	30.8	12.7	39.4
Foundation Floor Thickness	0.76	0.60	0.54	0.25	15.2	20.2	21.4	29.2	17.3	22.5	24.3	30.8
Glass Conductivity	0.83	0.03	0.76	0.07	11.8	35.2	17.6	35.5	14.5	34.4	17.2	33.6
Attic Insulation Conductivity	0.82	0.10	-0.14	-0.06	25.1	55.5	61.0	68.4	19.5	44.0	49.6	47.9
Lighting Energy Power Density	0.32	0.33	0.31	0.33	13.7	13.5	14.3	14.0	16.2	16.0	16.2	16.0
Equipment Energy Power Density	0.32	0.44	0.42	0.39	13.2	12.1	12.6	13.2	15.7	14.2	14.6	15.0
People Quantity	0.69	0.34	0.46	0.08	14.4	23.2	21.7	29.8	17.2	25.2	22.9	29.8
Ventilation Flow Rate	0.79	0.53	0.66	0.10	7.2	10.6	8.7	15.7	8.5	12.9	11.0	17.7
Infiltration Flow Rate	0.98	0.91	0.98	0.74	7.7	14.3	6.1	25.8	6.0	13.7	5.8	23.9
Maximum Heating Air Flow Rate	0.98	0.78	0.98	0.77	6.7	20.8	6.3	23.7	6.8	22.5	6.3	23.0
Range	0.00		1.00		0.0		70.0		0.0		50.0	

2.4.4 Decay-Curves Only

Results for when the ISM is trained on only decay curves are shown in Table 2.6. Expectedly, the 14-F scenario performs the strongest overall given that it contains the largest amount of data. Results are typically more affected with the decrease to 5 zones than with lowering the amount of decay curves. While the 5-H scenario is the most limiting, the prediction performance for several parameters (Lower Wall Insulation Conductivity, Roof Insulation Conductivity, Infiltration Flow Rate and Heating Air Flow Rate) remains high. As the decay curve process is more flexible than selecting a consistent day, week or month, it signifies that training with decay curves is more resilient to change.

Table 2.6 Error Metric Scores for decay curve results. *Performance is most comparable to the week results. Moving from 14 to 5 zones hindered performance more greatly than limiting the amount of decay curves.*

DECAY	R^2				MAPE				CV (RMSE)			
Scenario	14-F	14-H	5-F	5-H	14-F	14-H	5-F	5-H	14-F	14-H	5-F	5-H
Upper Wall Insulation Conductivity	0.83	0.72	0.58	0.02	27.7	34.6	40.3	42.2	19.9	25.3	31.2	47.5
Lower Wall Insulation Conductivity	0.97	0.95	0.97	0.93	8.1	11.8	9.1	14.2	8.3	10.3	8.0	12.7
Roof Insulation Conductivity	0.97	0.93	0.96	0.94	9.3	14.3	11.4	14.1	8.6	13.1	9.9	11.9
Foundation Floor Thickness	0.94	0.80	0.56	0.61	8.0	13.7	20.9	18.4	8.5	15.9	23.7	22.2
Glass Conductivity	0.91	0.72	0.66	0.70	8.9	18.7	17.0	19.3	10.5	18.4	20.2	19.1
Attic Insulation Conductivity	0.76	0.76	-1.54	-1.12	24.9	24.7	111.5	57.7	22.9	22.8	74.0	67.7
Lighting Energy Power Density	0.15	0.42	0.29	0.14	16.1	12.9	13.5	14.6	18.1	14.9	16.5	18.2
Equipment Energy Power Density	0.43	0.44	0.33	0.42	12.5	12.7	13.4	12.2	14.5	14.3	15.6	14.5
People Quantity	0.45	0.42	0.20	0.42	23.9	22.5	28.0	23.1	23.1	23.8	27.8	23.7
Ventilation Flow Rate	0.68	0.61	0.50	0.62	8.7	10.2	9.9	8.9	10.6	11.7	13.3	11.6
Infiltration Flow Rate	0.99	0.96	0.92	0.93	5.8	8.3	16.2	14.4	4.8	9.1	12.9	12.3
Maximum Heating Air Flow Rate	0.98	0.94	0.94	0.96	7.4	16.0	14.5	8.9	6.5	11.6	11.7	9.7
Range	0.00		1.00		0.0		70.0		0.0		50.0	

2.4.5 Variation in Prediction Performance

As indicated in the previous section, prediction performance can vary greatly for different parameters. Figure 2.5 illustrates the variation in R^2, MAPE and CV (RMSE) scores across all scenarios and parameters.

Figure 2.5 Variation in error metric scores for prediction performance across all scenarios and parameters. *Some very poor scores were omitted for clarity. Overall the month 14-C and 14-R scenarios performed the strongest.*

Aside from the Attic Insulation Conductivity, the prediction performance with the Upper Wall Insulation Conductivity has the largest variation overall. The month scenarios for 14-C and 14-R preform arguably the strongest overall with the week scenario for 14-C and decay curve scenario for 14-F performing solid as well. Ultimately, it can be observed that the Day 5-R scenario performs the worst with Day 14-R and Week 5-R also showing poor performance.

While there is a notable decrease in performance quality from moving from a month of data to a week, there's a significantly larger discrepancy in prediction performance when moving from week to day. While the method is able to accurately predict some parameters in most of the scenarios, caution would need to be applied if the parameter sought is non-influential as either a high quantity or quality (i.e., consistent) of data may be necessary. Results from training with decay curves are most comparable in performance to that of training on a week of data. This would suggest that training on decay curves is an appropriate alternative over the week scenarios especially ones that are randomized.

2.5 Conclusions and Future Work

This study examined the potential of a CNN based ISM for the purposes of predicting building characteristics. Once trained on multi-zone temperature time series inputs, prediction performance was examined across a variety of parameters for month, week and day durations. We also examined training on only instances in time for which internal temperature steadily decreased over a given period (decay curves). Several scenarios were introduced to examine the robustness of the prediction performance of the ISM including randomizing the time interval from which data was collected (not applicable when only trained on decay curves) and varying the number of zones for which data was provided.

The Infiltration Flow Rate and Heating Air Flow Rate parameters had the strongest performance for their predictions. While most other parameters had decent predictions in the best case (1 month of data for a consistent period and 14 zones), the accuracy could decrease substantially in other scenarios. Some parameters performed exceptionally worse when the amount of zones was reduced suggesting that predicting them accurately may be contingent on whether relevant zones for them are provided. The People Quantity and Ventilation Flow Rate

experienced unstable prediction performance outside of the 14-C scenarios. Removing the basement zone noticeably impacted their prediction performance, however some predictability still remained in 5-C scenarios suggesting that some determination could be made when the data was consistent. Future work could investigate this phenomenon more closely to determine exact causes of performance variability.

Randomizing the time interval usually had a more profound negative impact than reducing the number of zones. When trained only on decay curves, the model exhibited typically weaker performance compared to being trained on a consistent month of data, however, for most parameter predictions it outperformed a model trained on a month of data that was randomized.

Regarding future work, as training on only decay curves demonstrated higher robustness, training on different forms of data or different ISMs may prove more successful in this regard as well. One consideration would be to provide specific electrical consumption of different zones as this might pair better with some parameters such as Equipment Energy Power Density and Lighting Energy Power Density. Additionally, Recurrent Neural Networks are also developed to train on time series inputs, making them a potential candidate for ISMs.

For the purposes of retrofit planning, the method will now be applied with actual sensor data. This will demonstrate the suitability of the method in determining building characteristics. As noted in [19] however, training purely on synthetic data might be insufficient in being able to provide reasonable estimations, hence further efforts, such as transfer learning, might be necessary.

It would also be worthwhile to examine more complicated and realistic changes to the building energy model from which synthetic training data is derived. For example, ventilation and infiltration flow rates were treated as constant parameters in this study, however in reality they are continuously changing as a result of external variables such as wind. Modifying these parameters and others to produce a more accurate building energy model may be beneficial in this regard, but overall prediction accuracy may decrease as a result of the parameter values being more dynamic.

This paper has shown that inverse surrogate models based on CNN machine learning models have the potential to derive unknown building parameters from time-series temperature

data obtained from simulations. This is not possible with standard BES tools, which require a manual or computationally intensive calibration processes. In future work these methods will be further developed and applied to measured temperature data.

Chapter 3 Examining the Generalizability of Inverse Surrogate Models for Different Building Model Geometries and Locations.

3.1 Abstract

While building surrogate modelling has been shown to accurately replicate the outputs of computationally intensive building energy modelling, successfully adopting surrogate modelling in practice still has challenges. As surrogate models are machine learning models, they require an extensive quantity of training data in order to train effectively. The process of acquiring training data often requires numerous simulation runs of a building energy model. To offset this issue, surrogate models that demonstrate a suitable level of generalizability can be applied successfully to multiple projects without the need for the further generation of data.

This study examines the generalizability of multiple inverse surrogate models. Inverse surrogate modelling is a more difficult task than traditional surrogate modelling as it tries to extract building energy model inputs from output data. As the output data required to do this is often comprehensive, deep learning models are preferred. For the inverse surrogate models, a basic deep artificial neural network, convolutional neural network, recurrent neural network and transformer were examined. Output data in this study consisted primarily of temperature and energy time series data with input data being building energy model parameters reflective of thermally important building characteristics.

Generalizability is assessed by first training the inverse surrogate models on data from 3 separate building energy models. Each of the building energy models contain geometry that is randomly scaled. Additionally we examine training the inverse surrogate models on building energy model data produced with multiple locations as well as on data from all building energy models at once. Parameters relating to the building envelope demonstrated the highest prediction performance among the models, whereas the prediction performance for less influential parameters was more varied depending on the inverse surrogate model. Overall, the convolutional neural network typically outperformed the other models with the recurrent neural network and transformer producing slightly worse performance. The artificial neural network was unable to accurately predict parameters outside of a select few that were highly influential to

the time-series data. In the cases of training with data from multiple locations or all buildings at once, prediction performance decreased, however several parameters remained predictable.

3.2 Introduction

3.2.1 Background

Over the last few decades, Building Energy Modelling (BEM) techniques and methods have been rigorously applied to assist with early-stage building design, optimization, energy savings etc. [37]. With the Canadian Government funding new construction to address the needs of Canada's housing demand [38], the need for effective building performance simulation continues to grow. This is further exacerbated by the impact buildings have on Green House Gas (GHG) emissions. Globally, the continued growth of construction activities have led to increasing not only the GHG emissions from the act of constructing the buildings themselves, but have also contributing to the all-time highs of CO_2 emissions from building operational energy [39]. BEM software has been continuously developed throughout the last several years to further assist practitioners. Common BEM software programs include EnergyPlus [23], IES-VE [40], and DesignBuilder [41]. These computationally intensive software tools apply a complex set of inputs to develop a computerized model of the building from which a series of physics equations can be activated to calculate consumptions and emissions overtime.

BEM can be applied to various stages throughout the development of a building, including the preliminary design stage, developed design and post construction [42] [43] In the preliminary design stage process, building designers can rely on BEM to provide necessary insight into preliminary energy performance. While a comprehensive building energy simulation model can be considered too uncertain, due to the substantial number of known inputs required, additional BEM software has been introduced recently to aid decision making at this design stage [44]. Later in the design stage when more information of the building's design has been decided, a comprehensive energy model is more appropriate and practitioners can review the potential energy performance and determine if it is within desired limits. Once the post construction phase has been achieved, the use of digital twins becomes more apparent.

In each of these stages, the computationally intensive nature of BEM can be problematic. Running BEM simulations are often met with serious runtimes for many models, thereby hindering tasks which may necessitate multiple simulations. While simpler models alternatively used in the early design stage are more immune to high computational complexity problems, these models are more simplistic and lack some of the technical rigour that could be eventually desired. Furthermore, even at the design stage, some desirable input combinations (e.g., varied geometry combinations and different mechanical systems), may not be possible within the limitations of simpler models, thereby potentially weakening their usability. The issues of computationally intensive BEM can extend to later design stages as well, where finding suitable energy efficiency against increased costs can often require multiple simulation runs with varying inputs.

The performance of buildings after completion often do not align with the original models [43]. The common solution are calibrated building energy models adjusted such that their outputs are similar enough to those observed in reality. Creating a calibrated energy model is an iterative process for which input values are varied either manually or automatically until suitable outputs are acquired [45]. This iterative process involves the repeated use of simulations, for which the process can be beset again by long computational runtimes.

3.2.2 Surrogate Modelling

As a response to the hinderances that BEM currently face, an alternate method has emerged whereby a traditional BEM model is replaced with a building Surrogate Model (SM). A building SM is a Machine Learning (ML) model designed to replicate the performance of a computationally intensive BEM model while addressing the issue of long computational runtimes [12]. Instead of completing complex physics equations reminiscent of BEM, a trained building SM functions by placing accurate predictions on BEM output values almost instantaneously when provided relevant BEM inputs. Building SMs have demonstrated high accuracy in matching outputs from comprehensive BEM models [12]. Common examples include Artificial Neural Networks (ANNs) and Support Vector Machines. The creation of a SM can be organized into the following steps [46]:

1) Acquisition of training data.
2) Processing of training data

3) SM training
4) SM validation

ML models require a high volume of samples each composed of input and output data in order to train efficiently. For a building SM, input data would consist of numerical building properties referred to as parameters which can include the conductivity value of the wall insulation or the flow rate of infiltrating air. Output data would be calculated BEM outputs, such as energy consumption or internal temperatures. Each sample is composed through their own BEM simulation for which the input parameter values are applied to the actual BEM model (that the SM is trying to replicate) to obtain the corresponding output data. Acquiring enough training data in this fashion is a lengthy process, as it may consist of thousands of actual BEM simulation runs. While SM predictions are made near instantly, the process of acquiring the training data could be problematic. However, to counteract this, a SM can be considered generalizable if it can be applied to multiple projects or BEM models, thereby reducing the need to always create and train a new SM.

Once a sufficient quantity of training data has been obtained, it becomes necessary to perform data processing. Initially the integrity of the data should be assessed for missing data points or erroneous values. Fortunately, as the training data is generated via simulation, complete data integrity can usually be easily assured. Preprocessing in the form of data scaling should then be completed to prepare the data for training in addition to defining how to split the data into training, validation and test sets.

After training data has been obtained and appropriately processed, model construction and hyperparameter selection must be conducted prior to training. Model construction relates to the determination of which ML model, its components (e.g., layers) and the structuring of them. Hyperparameters are highly influential variables that relate to either the ML model's components or training ability. The ML model cannot learn its own hyperparameter values and it is ideal, though time consuming, to perform hyperparameter tuning by experimenting with different values and model configurations. It is noteworthy that an appropriate model construction and hyperparameter determination will not only influence prediction accuracy but also memory usage and training time among others.

After training and the generation of predictions, the SM is validated for its accuracy. This is often done through the use of statistical error metrics and common examples include the Coefficient of Determination (R^2), Mean Absolute Error (MAE) and the Root Mean Square Error (RMSE).

3.2.3 Surrogate Modelling with Deep Learning Models

With advancements in the ML domain, the rise of Deep Learning (DL) models and methods have provided recent SMs increased potential for successfully learning difficult tasks. DL includes a subset of ANN models that utilize multiple layers to continuously extract information. This is usually coupled with a more complex set of neurons allowing for advanced operations that provide higher performance within domains that contain large high-dimensional data, such as multiple time series, high resolution images, and complex text [47] Compared to shallow learning methods, DL methods have had less traction for BEM related tasks, however this has improved in recent years partly because of increasingly available powerful computing hardware [12] [48]. Common examples include multilayered ANNs, Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs). Sections 3.2.3.1 to 3.2.3.3 provide a brief description of their corresponding DL model as well as related studies.

3.2.3.1 Basic Deep Artificial Neural Networks

As suggested in their name, ANNs attempt to replicate the decision process of neurons within the central nervous system of biological animals [49]. Reminiscent of synapses in a brain, connections of artificial neurons within an ANN produce signals or values that are adjusted by a weight that is configured during the learning process. These neurons or nodes are collected in layers. Whereas early versions of ANNs utilized only one layer (shallow models), formulation of the back propagation algorithm enabled effective training of multi-layered ANNs [49].

With the rise in computing power and shear availability of data over the last several decades, DL ANNs have shown substantial success over other ML methods when tasked with large complex problems [50]. The main component of a basic ANN is the Fully Connected or Dense Layer. As shown in Figure 3.1, each node in the Dense layer forms a weighted connected to the each of the nodes in the previous layer, hence being “fully connected”.

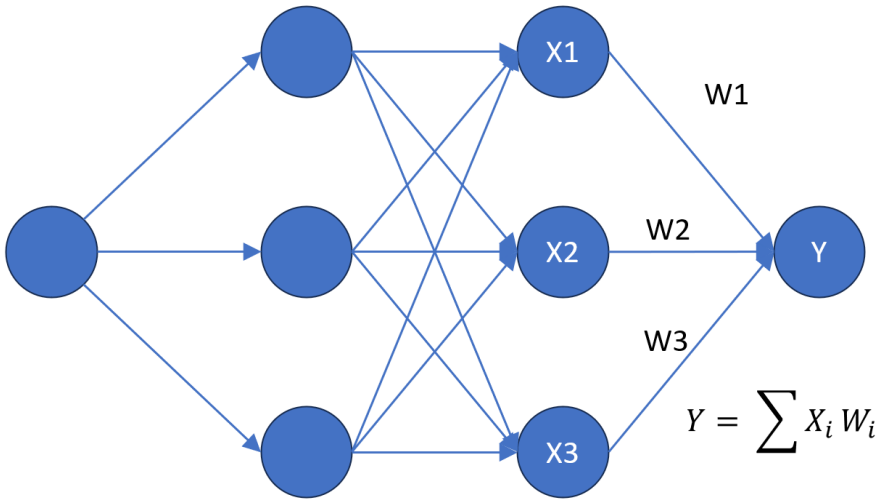


Figure 3.1 Simple ANN example. *This example features two dense layers in the middle (hidden layers).*

Olu-Ajayi, et al. examined the ability of a variety of ML models, including both deep and shallow ANNs, to predict annual energy consumption for different types of residential buildings [51]. Their study concluded that the deep learning ANN produced the best results overall, however, some other models were comparatively accurate and the building type was negligible on performance. Suryanarayana et al. compared DL to conventional ML methods for thermal load forecasting of district heating networks [52]. Their DL model consisted of a simple ANN with 2 hidden layers, while other models included a polynomial linear regression model and a ridge regression model. Their study examined two case studies involving different district heating networks in Sweden and in both studies, their DL model outperformed its simpler counterparts with a Mean Absolute Percentage Error (MAPE) of 8.08% and 4.15% respectively. Herbinger et al. developed a SM in the form of an ANN to calibrate a building energy model [14]. Their method is unique in that, once trained, the model uses itself to calibrate building parameter values via gradient descent. They compared their SM to a powerful ML optimizer in a controlled case study and found that their SM surpassed the ML in performance. They also compared the performance of the models in a real metered data case study and while it was more comparable, the SM was more consistent.

3.2.3.2 Recurrent Neural Networks

RNNs are a subset of ANNs that are specialized for learning information from sequential data. The significance of the model is that recurrent connections are placed between nodes in the layer such that information from the previous time step is sent forward, which enables the nodes at previous time steps to directly influence the output of future time steps [26]. This allows the model to learn sequential data more strongly compared to an ANN, as there is an immediate connection between each sequential value. This process is illustrated in Figure 3.2.

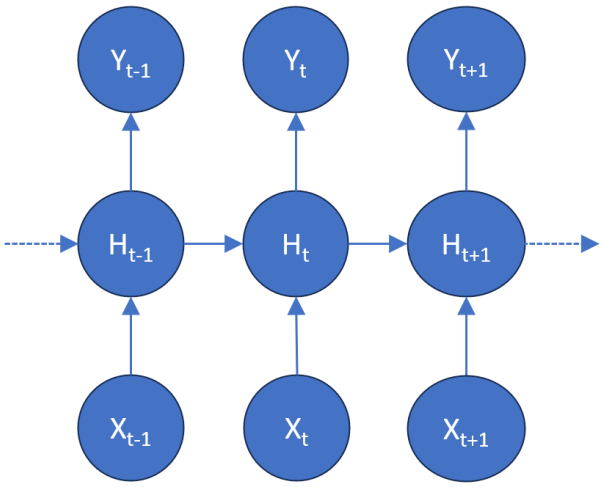


Figure 3.2 RNN configuration. *Inputs (X) are fed into the hidden nodes (H) which influence future outputs (Y).*

The major hinderance traditional RNNs face is the issue of the vanishing (or sometimes exploding) gradient. In cases of the vanishing gradient, long-term dependences become very difficult or time consuming for the model to learn as exponentially smaller weights are provided to long-term interactions [26]. As a response, different variants of RNNs, namely Long-Short Term Memory (LSTM) and Gated Recurrent Unit (GRU) models have seen prominence over traditional RNNs. LSTM models make use of a memory cell that modifies the RNN architecture by enabling gating signals to help handle challenges that can arise during model training [53].

Fan et al. examined multiple RNNs and input data strategies for short-term building energy predictions [54]. Their models included regular and bidirectional GRU and LSTM models as well as the possible inclusion of a 1D-convolutional layer beforehand. They note that a direct approach, where separate models are created for each time step, performed the best without significant computational burden and both the GRU and LSTM models provided better preservation of long-term temporal dependencies. They also noted that bidirectional operations

were beneficial for improving prediction accuracies for cooling loads. Jung et al. developed a multilayer attention-based GRU for the purposes of short-term load forecasting [55]. Making the GRU attention-based was chosen to help enable the model to understand which parts of the input sequence are the most important for load forecasting. When tested for three separate buildings they noted that their model outperformed others such as conventional LSTM and GRU models. Jang et al. employed three different LSTM models for the purposes of predicting heating energy consumption [56]. The first model had input data consisting of building environmental data, the second model also had building environmental data as well as outdoor environmental data and the third model had both sets of environmental data along with operation patten data. They found that the third model outperformed the other two while having the additional strength of better handling situations where the energy consumption experienced a sudden change.

3.2.3.3 Convolutional Neural Networks

Whereas RNNs are specialized to handle sequential data, CNNs are a particular type of ANN for the purpose of efficiently processing data with a grid-like topology [26]. Data in this form can include images or multiple concurrent time series. CNNs rely on a mathematical process known as convolution, which involves multiplying the input matrix by another, usually smaller, matrix referred to as the kernel matrix. The process is completed by multiplying the kernel matrix along portions of the input matrix (Figure 3.3). The unique property of the kernel matrix is that because it moves along the input, its weights are effectively shared. For this reason, CNNs are typically much more memory efficient than similar deep ANNs.

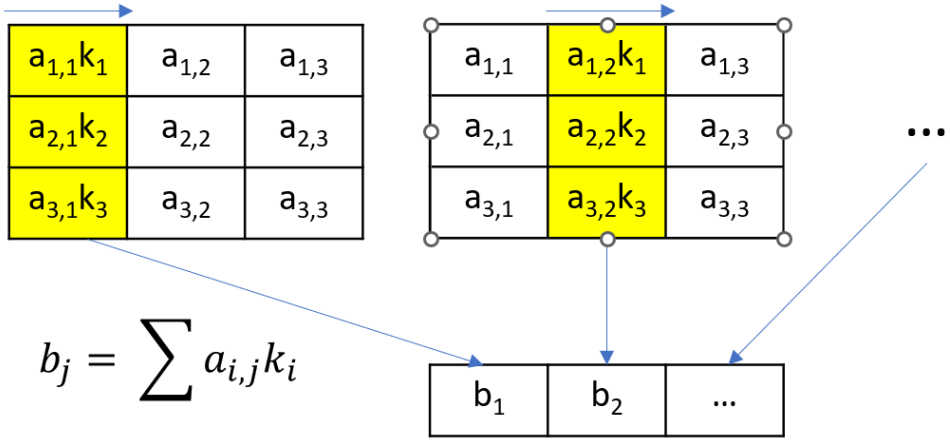


Figure 3.3 1-Dimensional Convolutional Process. *The kernel matrix is multiplied by sections of the input matrix to form the output. The kernel moves along during the convolutional process as indicated by the highlighted cells.*

Somu et al. incorporated k-means clustering into a hybrid CNN-LSTM model with the intention of building energy consumption forecasting [57]. The K-means clustering was performed to help organize the energy consumption trends, the CNN to extract features from non-linear interactions impacting energy consumption and the LSTM component to help control the long-term dependencies within their time-series data. When applied to a case study involving collected building energy consumption time series data, their hybrid model outperformed more conventional CNN and LSTM models. Westermann et al. developed a CNN based SM with the intention of predicting heating or cooling energy demand regardless of location [31]. The CNN processes weather-based time-series data before receiving additional inputs in the form of building parameters. When tested on unseen locations, the model experiences a MAPE of less than 3%.

3.2.4 Areas of improvement

While DL models provide practitioners increased opportunities for SM usage, they still face the same fault of requiring numerous amounts of training data. With growing popularity, it remains important that SMs express a suitable level of generalizability so that they can be readily applied to multiple projects without requiring the lengthy process of generating new training data each time. Generalizability relates to a ML model's ability to place accurate predictions on unseen data [26]. In the context of this study, we are focused on the generalizability that relates to multiple buildings (i.e. unseen geometry).

The intention of this study is to explore the generalizability of multiple DL models for building SM applications by examining the robustness of each model when trained on variable geometric BEM models. The BEM models used to create training data in this study are partly inspired by commercial office buildings and their basic geometry are randomly scaled to allow for a scenario where a SM needs to be used for another project. Additionally examined is the impact to prediction performance when trained on data from a BEM model simulated in multiple locations or being trained on data from all BEM models at once.

As deep learning methods exhibit their largest potential with high-dimensional data, this study examines the applicability of DL to inverse modelling. For BEM, inverse modelling is the notion of training a SM on BEM outputs to find inputs instead of the usual other way around. As the inverse task is being applied, a SM that trains and predicts this way can be referred to as an

Inverse Surrogate Model (ISM). This is often a much more difficult task as it is not entirely possible to map exact input values to relevant physics equations based on only output data as a different combination of values could result in the same output. Fortunately, many BEM outputs can be expressed as a time series and with a DL model's ability to extrapolate features by inferring the time series, obtaining individual values has a heighted possibility.

This paper compares an ANN, RNN, CNN and a transformer architecture. Both RNNs and CNNs were selected as they are well suited for an ISM trained on time series data, while a basic deep ANN helps provide a performance benchmark. In recent years, transformers have gained increased attention in machine learning communities for their ability to surpass RNNs and other ANN models for sequential text related problems [58]. Unlike RNNs, the transformer is able to model the relationship between all inputs in a sequence regardless of placement by applying a self-attention mechanism. This enables the model to understand the influence of elements more easily in the sequence on other potentially far away elements when compared to a RNN or CNN.

3.3 Methodology

The methodology of this study is broken into four sections: 3.3.1 training data organization, 3.3.2 geometric model design and development, 3.3.3 ISM model construction and 3.3.4 error metrics for evaluation. Figure 3.4 provides an overview of the methodology used.

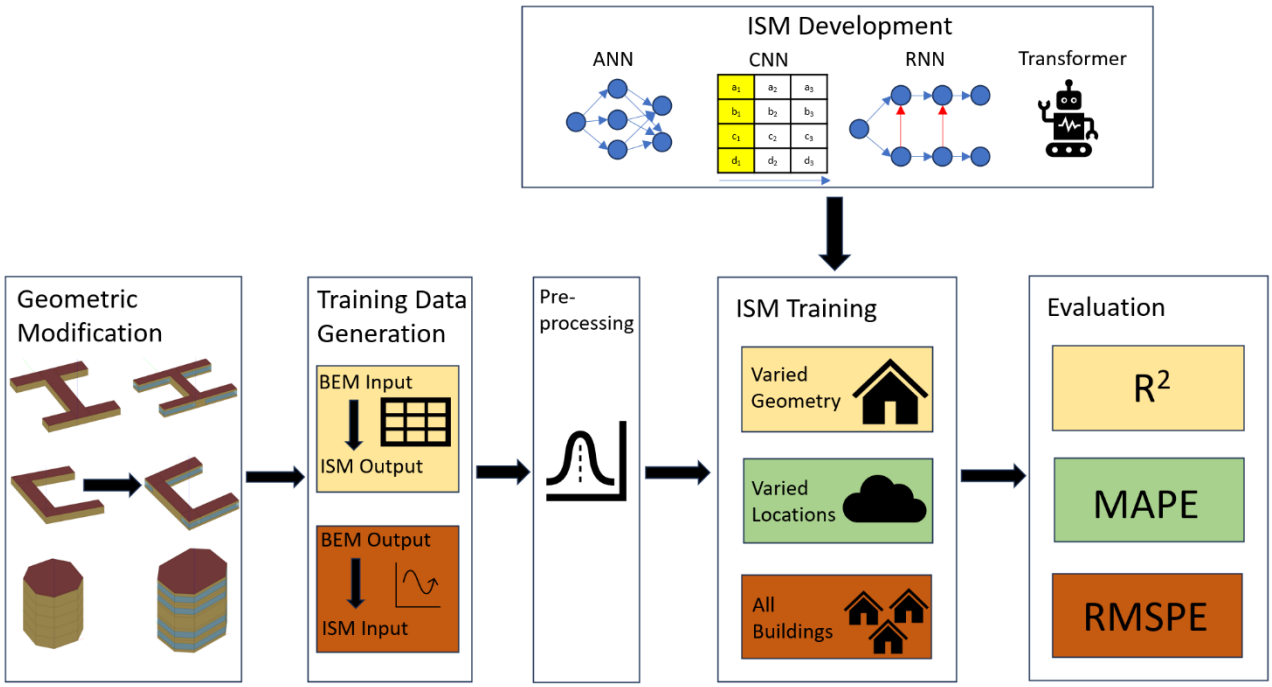


Figure 3.4 Study methodology. *Three different scalable core office shapes are used to produced training data. Prediction performance is examined with 4 different ISMs: ANN, CNN, RNN and a transformer. Prediction performance with each of the 3 core shapes is assessed as well as training for multiple locations and all shapes at once.*

Organizing training data relates to deciding on appropriate BEM model parameters and BEM output data to collect along with adjustments and necessary preprocessing. Geometric model construction examines the geometry of the BEM models themselves and other various aspects of them including scalability. ISM model construction relates to the determination of ISM layers and various hyper parameters. Model evaluation describes the various error metric equations used to evaluate the models in the results section of this study.

3.3.1 Training Data Organization

When training data is selected, it is imperative that there exists a strong connection between the inputs and the outputs. For the ISMs, this requires that the building parameters selected make a noticeable impact on the BEM outputs otherwise the ISM will struggle to learn the corresponding parameter values. This study prominently uses internal temperature time series data as an ISM input, as they are significantly affected by building envelope properties. Section 3.3.1.1 describes the parameter selection as they are the most influential on the chosen time series, which is then followed by describing the time series selected as well as their characteristics in section 3.3.1.2 and then how preprocessing is performed in section 3.3.1.3.

Lastly, we discuss the approach of acquiring training data when testing with multiple locations at once (section 3.3.1.4).

3.3.1.1 BEM Parameter Selection

The varied parameters used in this study are provided in Table 3.1 along with their ranges. The first eight parameters are typical BEM inputs that all have a varying impact on the internal temperatures of the BEM model. Predicting these parameters represents the primary task of the ISMs. The remaining four parameters are implemented with the intention of assisting with the development of suitable training data. Their values are additionally predicted by the ISMs, however, given their significant relevance to internal temperatures, the prediction performance on them is anticipated to be higher than the others.

Table 3.1 BEM Parameters and Ranges. *Geometric scale parameters are set as a percentage to retain consistency.*

Parameter	Ranges	Units
Wall Insulation Conductivity	0.01-0.1	W/mK
Roof Insulation Conductivity	0.01-0.1	W/mK
Glass Conductivity	0.005-0.03	W/mK
Lighting Energy Power Density	8-12	W/m^2
Equipment Energy Power Density	8-12	W/m^2
People Quantity	0.025-0.05	$People/m^2$
Ventilation Flow Rate	5-10	L/sPerson
Infiltration Flow Rate	0.1-1	L/sm2
Maximum Heating Air Flow Rate	25-250	L/s
Geometric Scale X Direction	0.01-1	
Geometric Scale Y Direction	0.01-1	
Geometric Scale Z Direction	0.01-1	

The ranges in Table 3.1 are based partly on those in [31] and selected to range from highly to moderately influential for internal temperatures. The conductivity parameters influence the overall effectiveness of the building envelope and their individual impact is additionally contributed by the shape modification of the structure. Increases in height lead to increases in wall and window surface area while the roof surface area remains constant, thereby causing an increase in the contribution of overall building envelope heat flow as a result of wall insulation and window conductivity.

Both the Lighting Energy Power Density and the Equipment Energy Power Density parameters influence the internal temperatures via internal gains from lighting and equipment respectively and as a result they are expected to have less impact than the building envelope parameters. Predicting these parameters will examine the ability to accurately discern smaller influences. For all BEM models, the setpoint schedule for the lighting was turned on fully between 5:00 AM to 8:00 PM and completely turned off otherwise. To differentiate it, the equipment energy setpoint schedule was turned on at all times. Differences in prediction performance between the two parameters would suggest that the ISMs would either be impacted or not with a varied setpoint schedule.

While the People Quantity parameter is similar to the lighting and equipment parameters, it is unique in that it directly influences the Ventilation Flow Rate. The Ventilation Flow Rate parameter represents the rate of the intended flow of air into the zone and is increased depending on the number of personnel. The Infiltration Flow Rate parameter is instead the unintended rate of airflow into a zone and is therefore unaffected by the number of personnel and is highly impactful on internal air temperatures.

Instances where the internal temperature changes are those which are the most valuable for the ISMs, which typically occur when the heating setpoint changes. When the temperature decreases rapidly, it can suggest that the combination of building envelope parameter values are poor for an energy efficient structure, while slow temperature changes during the same weather period would indicate a more energy efficient structure. The heating setpoint schedule begins at midnight with 16°C until 6:00 AM where it is increased to 20°C and then further increased to 23°C at 7:00 AM. At 9:00 PM the heating setpoint reverts back to 16°C.

The Ideal Air Loads system component in EnergyPlus is used to represent the heating supply. This component functions as a highly efficient Heating, Ventilation and Air Condition (HVAC) unit and can be used when developing a full HVAC model is not necessary [24]. The Maximum Heating Air Flow Rate parameter limits the ability of the component, which prevents the BEM model from reacting too rapidly to setpoint changes, therefore increasing ISM trainability.

As the different BEM models were modified by different absolute values, the geometric parameters were organized as a factor to keep consistency among them. Scaling by a value of 1 would apply the maximum increase, whereas 0 would apply the maximum decrease.

3.3.1.2 Time Series Composition

As the temperature within a BEM model fluctuates over time as a result of thermal setpoint changes and outdoor temperature, the parameters in Table 3.1 provide a strong influence on the temperature within the model. In this study, the time series selected were interior temperature for each zone (5 total), outdoor air drybulb temperature and the heating energy provided (Table 3.2).

Table 3.2 Time-Series ISM Inputs. *Cells highlighted in green represent parameters applied to ISMs in all scenarios. Those highlighted in blue are only applied to specific parts of the study.*

Time Series	
Zone 1 Internal Air Temperature	Solar Azimuth Angle
Zone 2 Internal Air Temperature	Solar Altitude Angle
Zone 3 Internal Air Temperature	Zone 1 Air System Sensible Heating Energy
Zone 4 Internal Air Temperature	Zone 2 Air System Sensible Heating Energy
Zone 5 Internal Air Temperature	Zone 3 Air System Sensible Heating Energy
Outdoor Air Drybulb Temperature	Zone 4 Air System Sensible Heating Energy
Heating Energy	Zone 5 Air System Sensible Heating Energy
Outdoor Air Wetbulb Temperature	District Cooling Energy
Diffuse Solar Radiation	Outdoor Air Relative Humidity
Direct Solar Radiation	Precipitation Depth

Instances of temperature decay when the heating setpoint is lowered are more important to the learning process than instances of temperature rise as during decay, the rate of temperature change is not influenced by heat being provided by the HVAC system. In this study, instances of

heat loss are referred to as decay curves while instances of temperature rise (usually between a lower heating setpoint and a higher heating setpoint) are referred to as rising curves. It is important to note that the Maximum Heating Flow Rate parameter remains useful even in situations of decay as heating would be applied once the new setpoint is acquired to prevent further loss of heat. Rising curves remain beneficial, however more unknown influences will occur. For example, while solar radiation may only play a small role during decay curves, as they will often happen at night, their presence will frequently be noticeable during rising curves which will more likely be present during daylight hours.

Training with each BEM model individually, as well as all together, was attempted with only decay curves. Training with a combination of decay and rising curves was done when training the ISMs on a BEM model for multiple locations. The intention was that the additional time series data would assist the ISMs in understanding location and help accurately predict parameter values. When training with a combination of decay and rising curves, the additional time series used are also listed in Table 3.2. Some of these time series are influential to the internal temperatures (such as Direct and Diffuse Solar Radiation).

Decay curve instances were computed at 10-minute intervals with 12 at a time (2 hours) so that the temperature decay could be adequately captured. A requirement that a difference of 0.05°C between each temperature value for the first 6 values was implemented so that the curves would not be too flat. This requirement was only examined on one selected zone in the BEM models with the assumption that the other zone temperature decay would be similar. As each sample contained a year of data, starting from the beginning of the year, a total of 84 suitable decay curves (1008 values for each time series) were extracted from each sample to resemble one week of training data.

When dealing with both decay and rising curves, the process was similar. Rising curves were also computed with the same minute long intervals and length and the same requirement of a 0.05°C buffer between the first 6 samples was applied. A combined total of 84 curves was again used, with rising curves being interwoven with the decay curve data.

3.3.1.3 Data Preprocessing

For each BEM model in this study, a total of 5000 samples were used to train the ISMs. An additional 5000 samples were used to train the ISMs in the case of varied location. With each

set of 5000 samples, 1000 samples were reserved for the test set and 800 samples for the validation set, leaving 3200 samples for the remainder of the training set. When training on data from each BEM at once, a total of 6000 samples (2000 from each) was used.

Prior to training any of the models, preprocessing was applied to the data in the form of normalization. Preprocessing data such that it resembles standard normally distributed data is commonly performed on ML models as poor training ability can result otherwise [28]. Given that the data in this study consists of various timeseries, the mean and standard deviation was computed across all samples in the training set for each individual time series. These values were then used to normalize both the training and test set with the test set using the training set mean values and standard deviations. Preprocessing was only conducted for training purposes. When model predictions were acquired, values were then unnormalized prior to error metric calculations.

3.3.1.4 Varied Locations

Regarding locations, the city of Victoria, Canada was the only city used to create training data with each BEM model. When investigating ISM performance when trained on data from a single BEM model in multiple locations, only North American cities were examined, which offered a suitable selection of varying climates. Varying location was simply completed by using a different weather file for each location. While determining location is a categorical problem instead of a regression problem, to keep consistency, the ML models predicted latitude and longitude values. The cities chosen are provided in Table 3.3.

Table 3.3 Locations used for varied weather data and their climate zone [59].

Location	ASHRAE Climate Zone
Victoria, BC, CAN	4C
Edmonton, AB, CAN	7
Winnipeg, MB, CAN	7
Toronto, ON, CAN	5A
Anchorage, AK, USA	7
Los Angeles, CA, USA	3C
Denver, CO, USA	5B
Miami, FL, USA	1A
Las Vegas, NV, USA	2B
Austin, TX, USA	2A

3.3.2 Geometric Model Design and Development

Each of the three BEM models were based on commercial office building configurations (Figure 3.5). As only the overall dimensions of the BEM models were modified in the model (length, width and height), we separately created different core shapes so that it could be understood whether an ISM is universally strong as a parameter predictor for multiple shapes or only specific ones. Each BEM model was composed of a total of five thermal zones, between which conduction occurs during simulations and temperature time series data is collected. Windows were included with a Window to Wall Ratio (WWR) of 50%. The WWR was held constant regardless of the change in wall size (i.e., windows would scale along with the wall). Windows were omitted from one zone (Z3) in each model to help the model differentiate the impact of them in regards to other parameter influences.

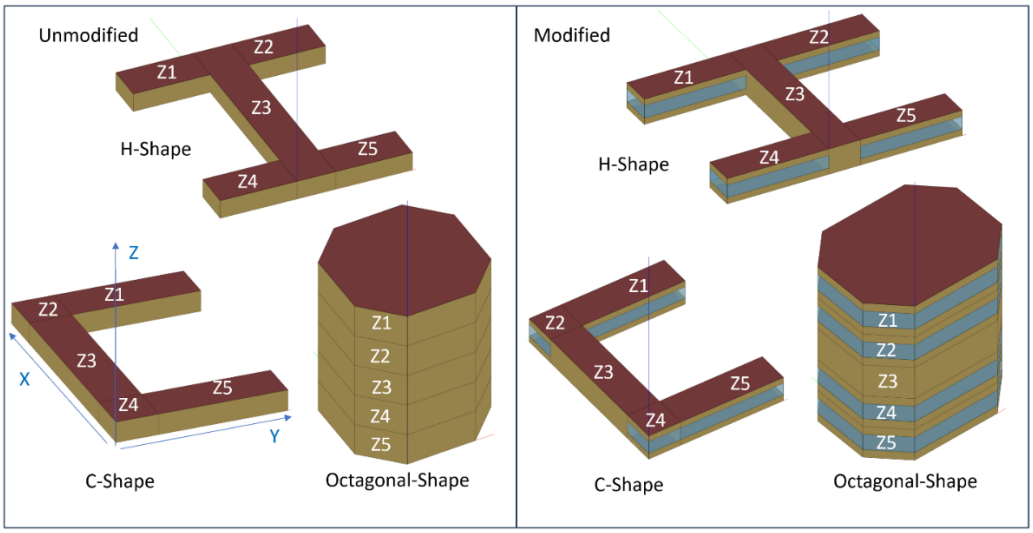


Figure 3.5 BEM geometries. *Zones are labeled Z1 through Z5. Windows were omitted from Z3 in each model to provide variation between the zones..*

The H-Shape model serves as the benchmark for comparisons, as each of its zones are scaled accordingly. The C-Shape model retains large similarity to the H-Shape, however scaling is reduced as Z2 and Z4 zones only shift along to accommodate scaling of the other zones and Z1, Z3 and Z5 only increase or decrease in their longer dimensions. The Octagonal-Shape model, differs from the others as it is the only model with multiple stories and nonrectangular zones. When the height is scaled, the height of each zone is modified, thereby significantly affecting the overall height of the structure. As the models are scaled differently, it may prove insightful in whether the degree of scaling affects predictions.

3.3.3 ISM Model Construction

While the ISMs have fundamental differences between each other, a learning rate of 0.001, a batch size of 64 and a choice of 150 epochs was used during training for each model. Furthermore, the Adam optimization and loss function of Mean Square Error (MSE) were implemented for each ISM.

Each model features some dense layers and dropout layers. Dropout layers are a computationally inexpensive method used to prevent overfitting by temporarily eliminating nodes during training [26]. Overfitting occurs when the ML model picks up noise that only exists in the training set during the training process, which leads to over-promising results. Along with the dropout layers, overfitting was prevented by Early Stopping, that would stop training prematurely if it was observed that the validation loss was no longer decreasing.

Hyperparameter values were chosen based on a similar study [31] as well as past experience. It is possible that different ISMs may benefit from different hypermeter considerations, which would incentivize hyperparameter tuning. Given the number and complexity of the ISMs used, hyperparameter tuning and varying of ISM model constructions would be complicated and is left for future work.

3.3.3.1 Artificial Neural Network

As the ANN serves as the benchmark for comparisons between other ISMs, it was decided to keep the ANN simple. The combination of Dense and Dropout layers that forms the model is illustrated in Figure 3.6. As the Dense layers are ignorant of the sequential nature of the data, inputs were simply formatted into a single row and fed into the model 1-dimensionally. The number of nodes contained in each hidden layer were twice the length (1008) with the exception of the last Dense layer before the output which was a quarter of this (252).

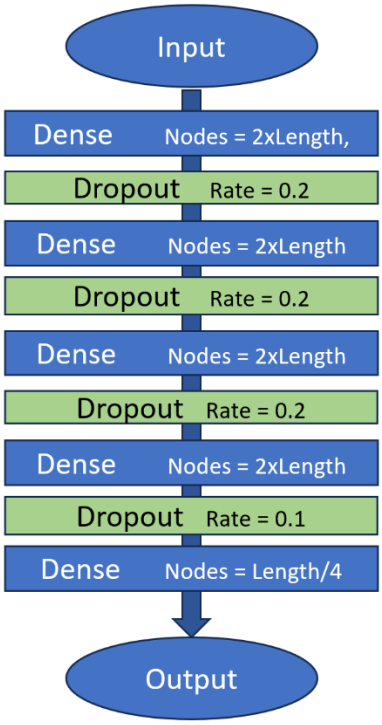


Figure 3.6 ANN Construction. *Dense layers are alternated with dropout layers.*

Aside from the final Dense layer, each hidden Dense layer uses the Rectified Linear Unit (ReLU) activation function. This remains the same for the other ISMs. An activation function is commonly applied to layer outputs to provide non-linearity. The Rectified Linear Unit (ReLU) is one of the most common activation functions for ANNs whereby values less than 0 are instead replaced with 0 and values greater than 0 remain unaffected as shown in equation 3.1 [26].

$$y = MAX\ (0, X) \quad (3.1)$$

3.3.3.2 Recurrent Neural Network

Compared to the ANN, the RNN in this study employes three LSTM layers early in the model to help with the sequential inputs. As shown in Figure 3.7, interwoven between the LSTM layers are dropout layers followed by a series of dense and dropout layers. 2-dimensionality is retained throughout the model until the global average pooling layer near the end of the model converts the data to 1-dimension. After another dense layer, outputs are produced by the model.

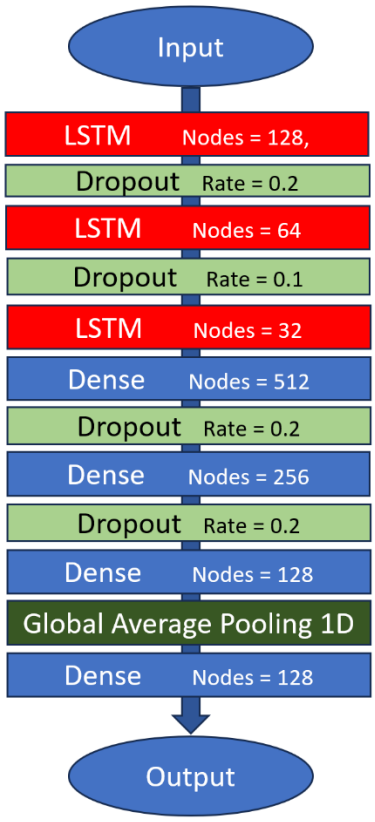


Figure 3.7 RNN construction. *This model features 3 LSTM layers to provide greater training with the time series sequential data. The pooling layer at the end is implemented to convert the data to 1-dimension.*

3.3.3.3 Convolutional Neural Network

The CNN used in this study is partly based on the model used in [31]. As shown in Figure 3.8, the CNN is formulated into two main blocks of several convolutional and assisting layers. Following each convolution layer is a batch normalization layer that is provided before the ReLU activation. Batch normalization provides reparameterization within a network to help benefit updates as they occur across multiple layers during training [26]. To help prevent the occurrence of the vanishing gradient issue, residual connections were provided at the end of each block creating another connection to the input preceding them. As with the RNN, a global average pooling layer is provided near the end of the model to transfer the data from 2-dimensional to 1-dimensional.

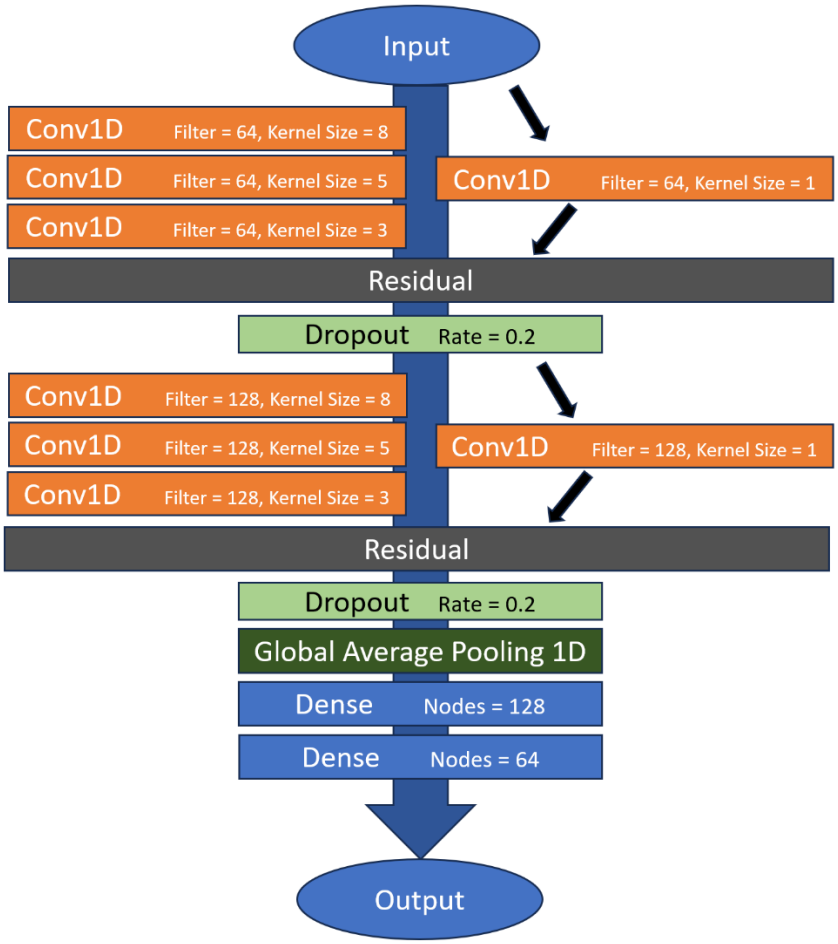


Figure 3.8 CNN construction. *In two occasions, three convolutional layers with decreasing kernel sizes are computed in parallel with a single convolutional layer with a kernel size of 1 which are then connected with a residual layer.*

1-dimensional convolution was used throughout this model. Compared to its 2-dimensional counterpart, 1-dimensional CNNs have substantially less computational complexity, which makes them well-suited for lower cost tasks involving 1-dimensional signals [30].

3.3.3.4 Transformer

The attention mechanism employed by transformer models can be viewed as a mapping of an input query to a dictionary of key-value pair to produce an output which is formed as a weighted summation of the values [60]. In this study, we utilize global self-attention for our time series inputs. Global self-attention involves feeding the model inputs as both the query and key-value pairs (Figure 3.9), thus enabling values throughout the time series to be influential [58]. This effectively allows each element in the input to interact with each other element directly with all the outputs being computed in parallel.

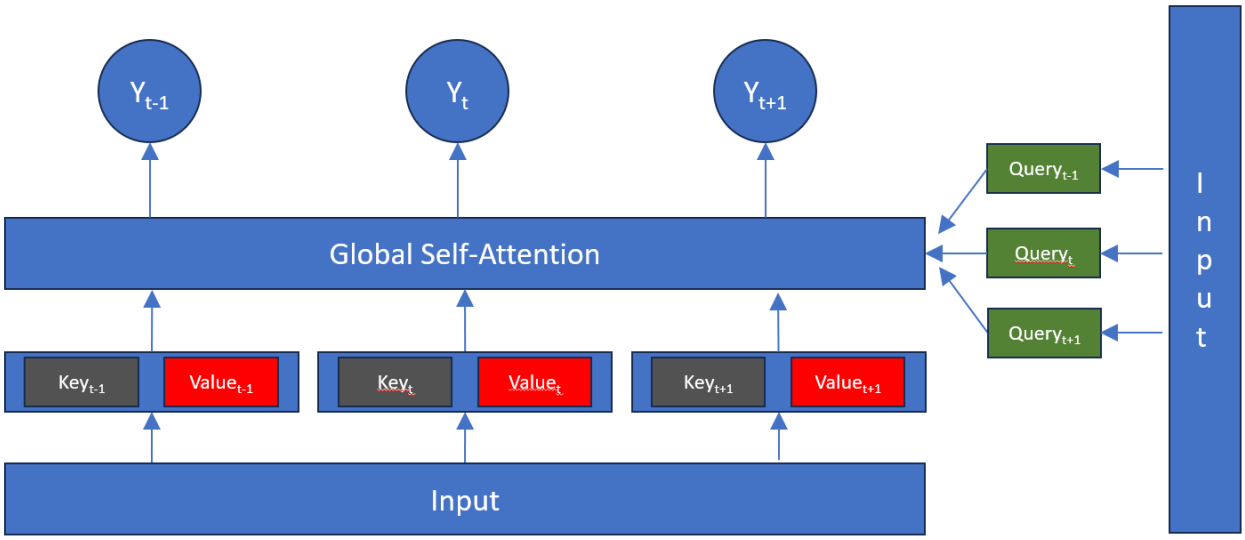


Figure 3.9 Global Self-Attention mechanism. *The query and key-value pairs are provided to the global self-attention layer. In our case this is a Multi-head Attention layer.*

The transformer architecture used in this study is provided in Figure 3.10. The model includes several multi-head attention layers followed by a residual connection to its input and then a linear normalization layer. Its structure is partly based on [60]. Multi-head attention allows for multiple representations of the inputs into the attention mechanism, which enables it to learn different features from each input [58]. In this study, to keep computational cost low during training, we used 2 heads with a size of 128. Similar to batch normalization in the CNN, the linear normalization layer helps maintain trainability throughout the learning process by helping maintain normalized values throughout the model. A small one-dimensional CNN block is applied to the end this model to help manage the sizeable data still present after the final attention block. As with the CNN and RNN before, Global Average Pooling was then applied.

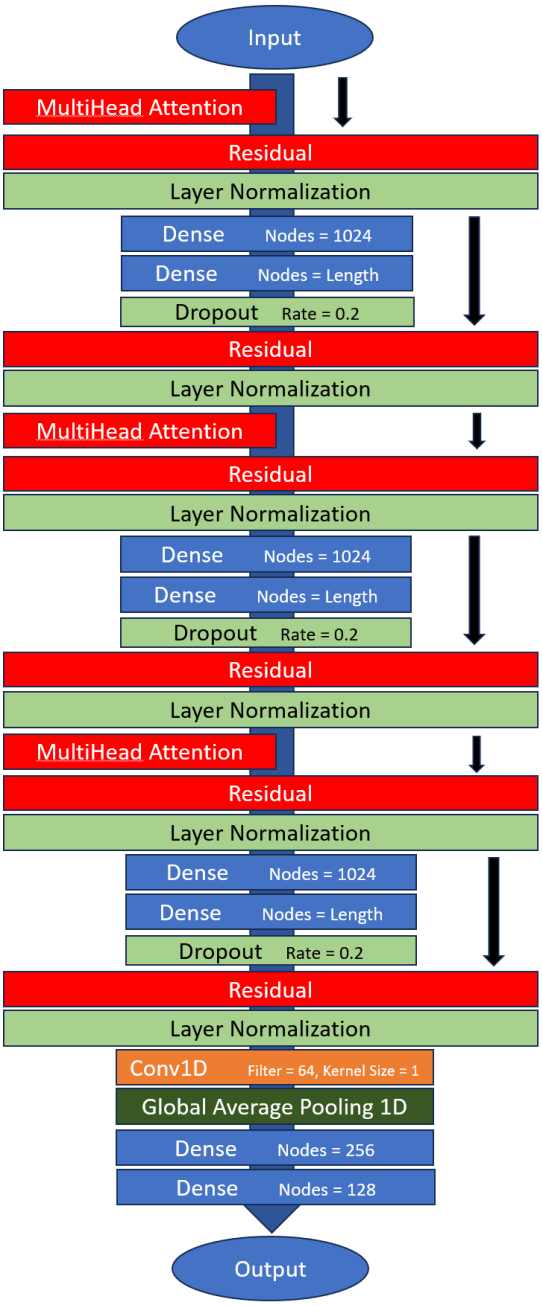


Figure 3.10 Transformer construction. *Arrows clarify a connection between layers. A single convolutional layer is provided near the end of the model to assist with the lengthy data still present.*

3.3.4 Error Metrics for Evaluation

Error metrics provide a statistical means of inferring ML model performance. These equations receive pairings of ML predictions and their corresponding correct value as input. By including multiple error metrics, a deeper understanding of model performance can be obtained, as some models may perform better or worse across different metrics. This study examines performance

with three different error metrics: the Coefficient of Determination (R^2), Mean Absolute Percentage Error (MAPE) and the Root Mean Square Percentage Error (RMSPE):

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \widehat{y_i})^2}{\sum_{i=1}^{n}(y_i - \overline{y_i})^2} \quad (3.2)$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\frac{|y_i - \widehat{y_i}\,|}{y_i} \quad (3.3)$$

$$RMSPE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\frac{y_i - \widehat{y_i}}{y_i}\right)^2} \quad (3.4)$$

The three variables in the above equations, y_i, $\widehat{y_i}$, $\overline{y_i}$ represent the actual values, the predicted values and the mean of the actual values respectively.

The Coefficient of Determination provides an assessment of how well the predicted values fit the actual values. The R^2 value typically ranges from 0 to 1, where a value of 1 indicates a perfect fit. As the error metric represents the degree of fit of the model, one of its benefits is the ease of comparing scores across multiple studies [33].

MAE and RMSE are statistical error metrics extensively used for evaluating model performance. There are various arguments and discussions on the applicability of the RMSE and MAE in comparison to each other and which to use [61][62][34] and therefore both are considered in this study. Referred to as relative or percentage errors, MAPE and RMSPE are preferred in this study over their absolute error equivalents (MAE and RMSE) as the parameter values being predicted have different units and scales. Expressing errors as a percentage helps ensure easy comparison between them.

MAPE is commonly used when relative errors are needed; reasons include its intuitive interpretation and adaptability for forecasting applications [63]. The quadratic term in the RMSPE equation helps emphasize instances where significant differences between values exist. In comparisons between models, situations where MAPE scores are similar, but RMSPE scores are farther apart, may signify that a higher percentage of outliers are present in the predictions of one model compared to the other.

3.4 Results

The results of this study are presented in three sections. Firstly, the performance of the ISMs are examined on each of the three geometric models individually. By doing so, we can assess which ISMs are the best performing and if performance greatly differs between geometric models. The second section of the results examines the prediction ability of the ISMs when trained on BEM data from multiple locations. Also included is a comparison between training with only decay curves from the first section and training with decay and rising curves. The last section of the results examines the performance of the ISMs when trained on data from all the BEM models at once. Only one location is used as it was expected that the decrease in ISM performance would already be significant.

3.4.1 H-Shape Model Results

Table 3.4 gives the error metric results for the ISM models when trained on data from the H-Shape BEM model. Overall, the CNN outperforms the other models with the RNN and transformer exhibiting similar performance to each other. Unsurprisingly, while the ANN is able to have decent prediction performance with a few parameters, it is severely lacking in performance compared to the other ISMs. This shows that the ANN is often unable to discern the influence of less impactful parameters.

Table 3.4 H-Shape results table. *The CNN performs the strongest overall while the RNN and transformer (TRA) are closer. The ANN expectedly underperforms.*

H-Shape	R^2				MAPE				RMSPE			
Scenario	ANN	CNN	RNN	TRA	ANN	CNN	RNN	TRA	ANN	CNN	RNN	TRA
Wall Insulation Conductivity	0.30	0.84	0.85	0.55	49.2	18.8	18.4	40.8	81.3	28.0	34.8	58.2
Roof Insulation Conductivity	0.35	0.91	0.83	0.21	49.6	15.2	19.8	47.9	84.9	25.0	33.3	61.9
Glass Conductivity	0.58	0.96	0.93	0.93	27.0	7.2	9.8	9.1	39.2	10.2	14.9	14.3
Lighting Energy Power Density	0.11	0.93	0.69	0.63	9.6	2.4	5.0	5.6	11.6	3.2	6.4	7.2
Equipment Energy Power Density	-0.02	0.55	0.62	0.55	10.1	6.8	5.7	6.2	12.1	7.5	7.4	7.5
People Quantity	0.02	0.32	-0.03	0.24	28.8	24.3	30.6	23.7	36.9	33.9	41.0	30.7
Ventilation Flow Rate	0.25	0.72	0.49	0.29	14.9	8.5	11.2	14.6	18.9	11.0	14.7	18.8
Infiltration Flow Rate	0.89	0.98	0.98	0.98	18.1	7.9	6.5	7.8	29.2	13.6	10.3	11.9
Maximum Heating Flow Rate	0.91	0.99	0.98	0.93	16.0	4.3	6.8	13.2	25.6	6.9	10.4	17.5
Scale - X	0.34	0.95	0.94	0.91	>100	28.6	28.5	28.2	>100	83.6	86.6	59.1
Scale - Y	0.06	0.83	0.72	0.81	>100	43.0	54.6	43.9	>100	>100	>100	>100
Scale - Z	0.69	0.97	0.96	0.94	75.6	17.3	19.5	20.6	>100	50.0	57.2	45.9
Range	0.00		1.00		0.0		70.0		0.0		50.0	

Based on the R^2 error metric, predictions for the Glass Conductivity, Infiltration Flow Rate and Maximum Heating Flow Rate parameters preform the most consistently strong. Aside from the Maximum Heating Flow Rate, this can be attributed to their involvement as a

component of the building envelope and thereby having a significant impact on the internal temperature of the structure. The Wall Insulation Conductivity and Roof Insulation Conductivity remains highly predictable for both the CNN and RNN, though the performance with the transformer is significantly worse. It is unclear why this occurs, however, as the parameters influence surfaces of every zone, the transformer may struggle to separate the influence of it.

While the R^2 values for the Lighting Energy Power Density and Equipment Energy Power Density are lower than some of the more predictable parameters for the RNN and transformer, the error for the other metrics is among the lowest. This is the same for Equipment Energy Power Density for the CNN. This would suggest that while a goodness of fit is not as achievable as some of the other metrics, their predictions remain relatively consistent (i.e., lack of outliers). Quantile-quantile plots are provided in Figure 3.11 to help illustrate this. This also remains especially true for the ANN, as the R^2 scores suggest practically no fit, however both the MAPE and RMSPE remain low.

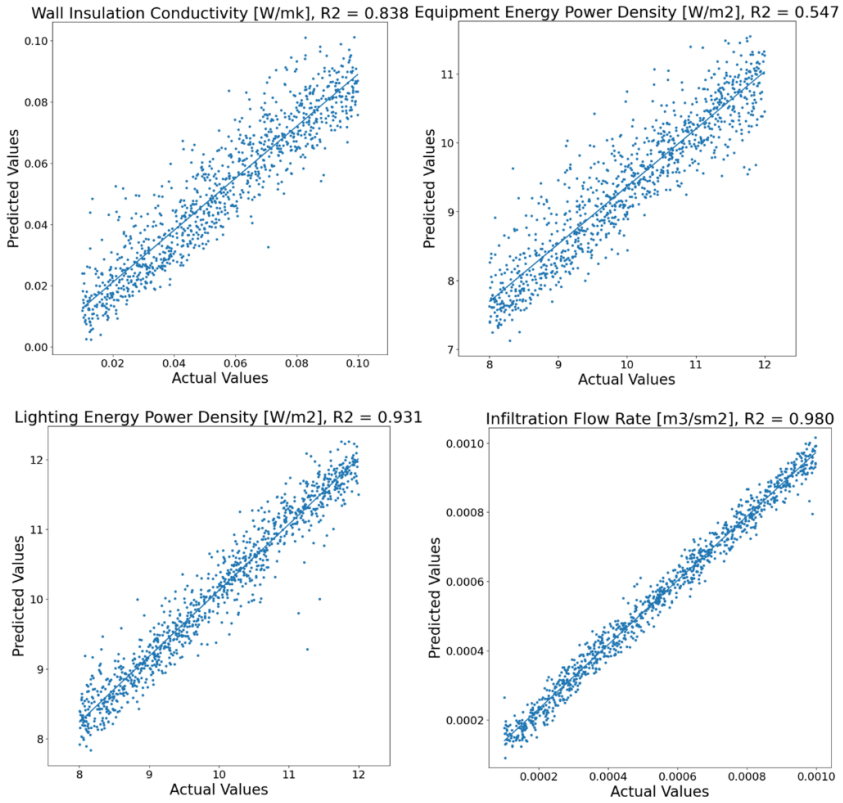


Figure 3.11 Quantile-Quantile plots for select parameters from the CNN. *Parameter predictions with a smaller R^2 have a less linear shape.*

Additionally, poor prediction performance on the Ventilation Flow Rate parameter is partly a result of the low performance for the People Density parameter. As the Ventilation Flow Rate is tied to the People Density, making accurate predictions remains difficult as the ISMs struggle to learn the influence of the quantity of people.

Another notable aspect of the results provided in Table 3.4 is how significantly large the RMSPE are for each of the predictions on the geometric parameters, even though the R^2 remains relatively solid throughout (ignoring the ANN). Given that there is also a large discrepancy between the RMSPE and MAPE scores, this implies that significant outliers have occurred (Figure 3.12).

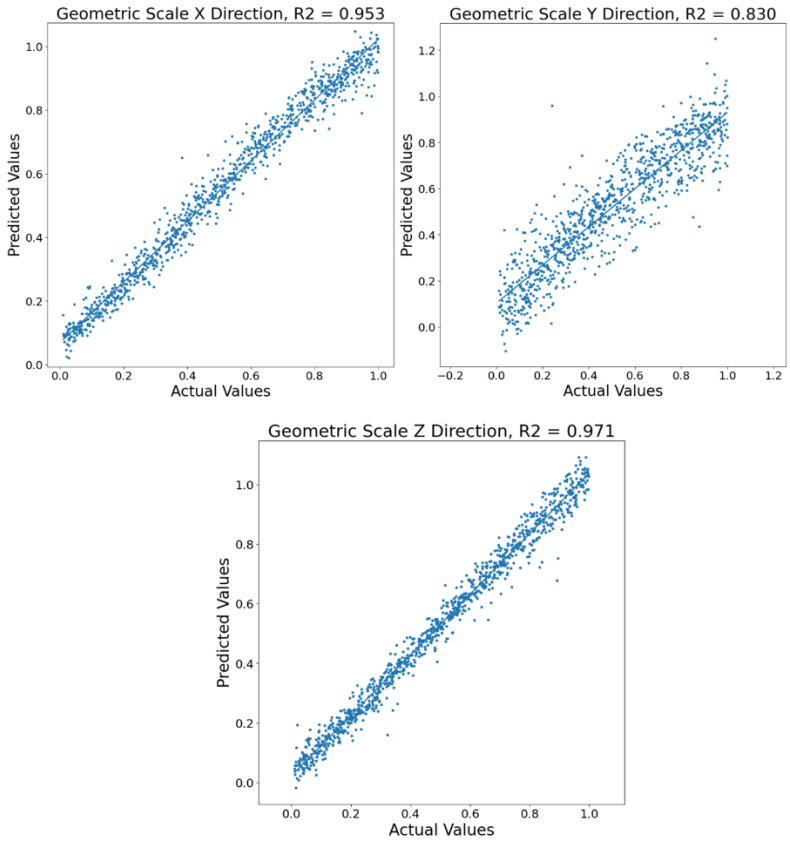


Figure 3.12 Quantile-quantile plots for the geometric parameters from the CNN. *Note that the Y scaling still appears linear though it often encounters more significant outliers.*

Lastly, Figure 3.13 demonstrates the decrease in loss over the number of epochs. As the model is trained to predict all parameters, the loss value is influenced by the more difficult parameter values to predict. It can be observed that while the ANN experiences an immediate sudden drop in loss, it plateaus higher and more rapidly than the other models. No more values

are reported when the Early Stopping is reached. Due to their similar structures, both the ANN and RNN exhibit less fluctuations than the CNN and transformer. Fluctuations with the transformer possibly help explain the poor prediction performance with some parameters (e.g. Roof Insulation Conductivity) as the consistency in making accurate predictions varies. As neither the CNN, RNN or transformer finish before 150 epochs, the number of training epochs could be increased at the consequence of longer training time.

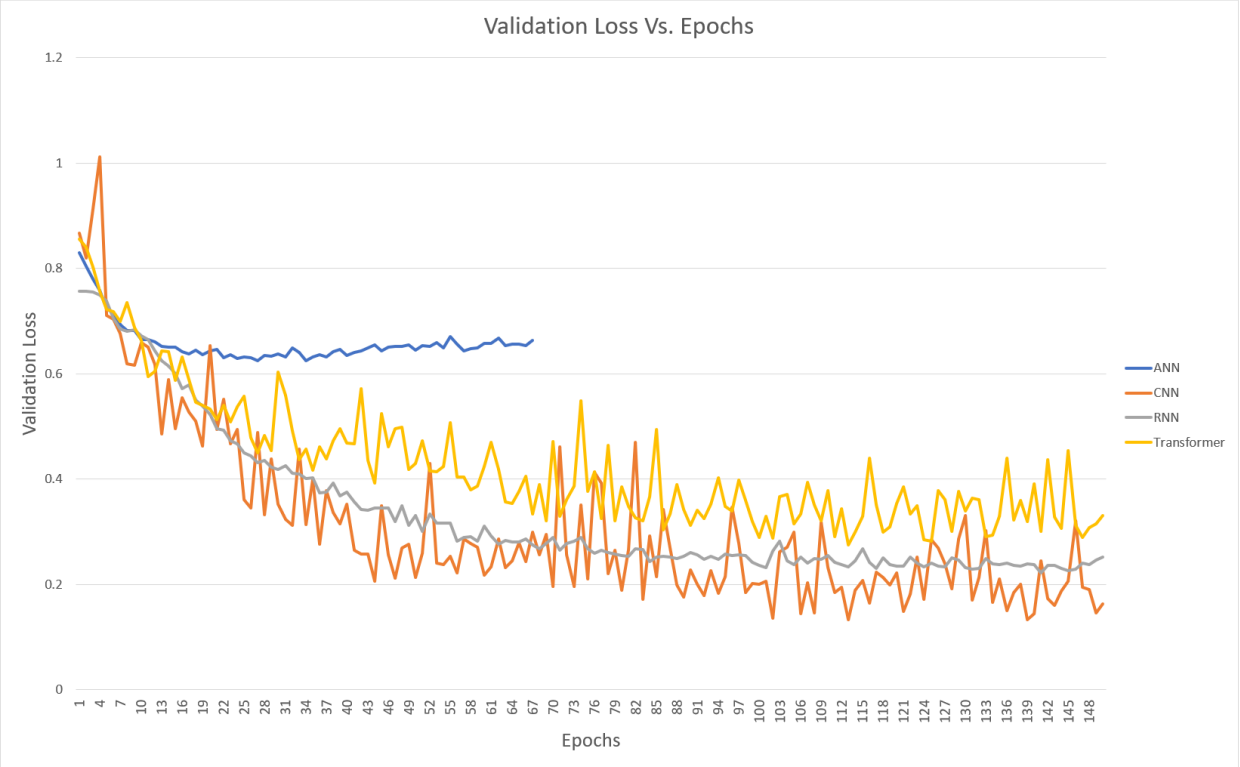


Figure 3.13 Validation Loss vs. Epochs. *The CNN and Transformer experience greater fluctuations suggesting training is inconsistent. The ANN is stopped early to prevent overfitting.*

3.4.2 H-Shape, C-Shape and Octagonal-Shape Result Comparisons

R^2 error metric comparison results when the ISMs are trained on data from each of the BEM models are shown in Figure 3.14. Tables and figures for the other error metrics are provided in Appendix A, along with those for sections 3.4.3 and 3.4.4. Findings with MAPE and RMSPE are similar to those in Table 3.4. Metric values for the geometric parameters have also been omitted for clarity; overall findings are again similar (i.e. high R^2, MAPE and RMSPE scores).

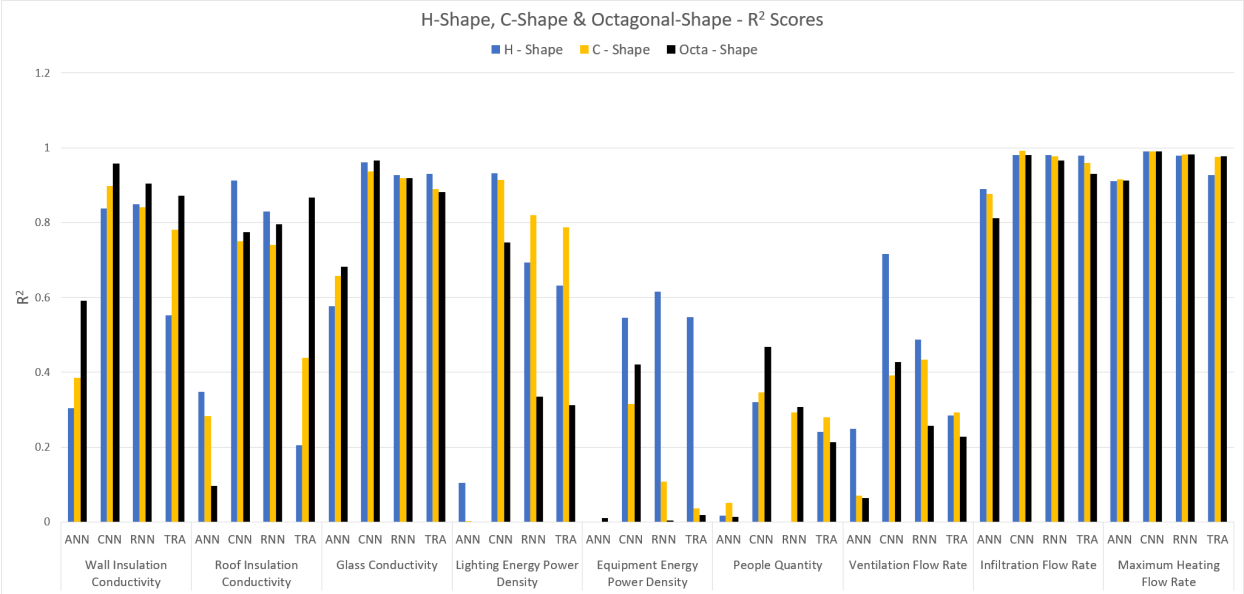


Figure 3.14 H-Shape, C-Shape and Octagonal-Shape - R^2 Scores*. Prediction performance remains similar between the models.*

It is observed that switching to a different BEM model for training data does not significantly impact overall performance for the ISMs. An exception to this is the transformer has much stronger predictions for the Wall Insulation Conductivity and Roof Insulation Conductivity with the data from the other BEM models. Overall performance among the ISMs remains similar with the CNN outperforming both the RNN and transformer models, while the ANN poorly performs on all parameters aside from the Maximum Heating Flow Rate and Infiltration Flow Rate.

One of the more notable differences between the BEM models is the higher prediction performance the ISMs have with the H-Shape model data for the Equipment Energy Power Density. Unlike similar parameters (Lighting Energy Power Density and People Quantity), the equipment parameter has a fixed schedule as it remains active overnight. The higher R^2 score indicates that there exists more instances in the H-Shape model data where its influence is noticeable. Interestingly, the Lighting Energy Power Density parameter has noticeably higher predictability for the C-Shape (aside from the CNN), which would suggest that the lighting being active is more frequent. As the equipment parameter would be less isolated in this dataset, it would be expected to be harder to predict. Both these parameters are poorly predicted with data

from the Octagonal-Shape model. As the floor area remains the same for each zone when scaled, these parameters have no difference in heat provided between zones, making it so that the ISMs have little ability to understand the impact of geometry scaling on them.

3.4.3 Decay and Rising Curves and Variable Locations

Unlike the results provided in Figure 3.14, when the ISMs are provided training data with decay and rising curves or from BEM simulations with variable locations, performance differs substantially (Figure 3.15). Overall, the ISMs trained on a combination of decay and rising curves exhibit superior performance compared to only being trained on decay curves. The performance improvement can be attributed both to providing the rising curves as well as the overall increase in time series for each sample. Unfortunately, there is a decrease in prediction performance when the ISMs are provided BEM data produced with multiple locations, as is to be expected.

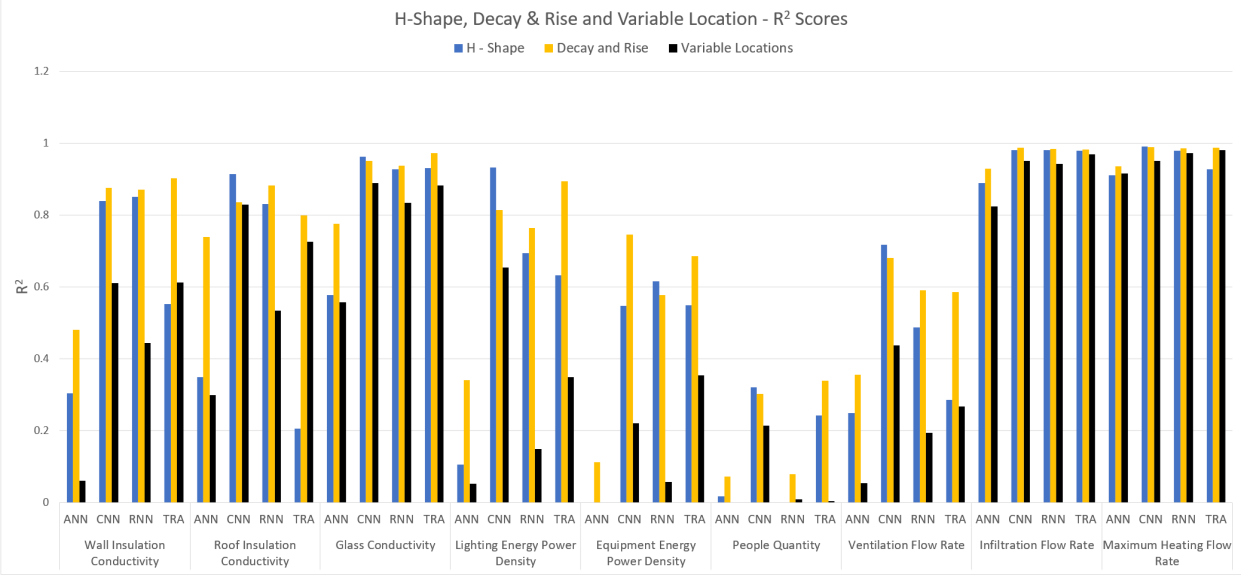


Figure 3.15 H-Shape, Decay & Rise, Variable Location - R^2 scores. *When trained with decay and rising curves, the Transformer achieved notably stronger prediction performance. All models performed expectedly worse with variable locations.*

Both the ANN and transformer ISMs experience the largest increase in prediction performance when switched to a combination of decay and rising curves. This shows that the models can take significant advantage of the increase in time series provided, which is especially true in the case of the transformer as it outperforms the CNN in some instances. The ANN still performs poorer than all other ISMs. The CNN performance decreases slightly for several

parameters suggesting that increasing the amount of data may be prohibitive for it to learn from, or that it benefits more greatly from only using decay curves.

Whereas in Figure 3.14 where both the People Density and the Ventilation Flow Rate parameters were similar in terms of performance, this is less apparent when the ISMs are trained on a combination of decay and rising curves. While the performance for the People Density parameter decreased, it increased for the Ventilation Flow Rate. With rising curves, ventilation would become notable as a countermeasure to increases in heat, when scheduled occupancy occurs. With rising curves, the small portion of heat provided by the quantity of people could get lost with the rest of the heat sources.

3.4.4 All Buildings

The results in Figure 3.16 show the R^2 score comparison between the ISMs being only trained on the H-Shape model compared to being trained on data from each BEM model at once. Expectedly, the overall prediction performance decreases for practically all parameters. Performance decreases are more significant on parameters that previously observed more varied prediction performance in Figures 3.14 and 3.15 (i.e. Roof Insulation Conductivity, Lighting Energy Power Density and Equipment Energy Power Density).

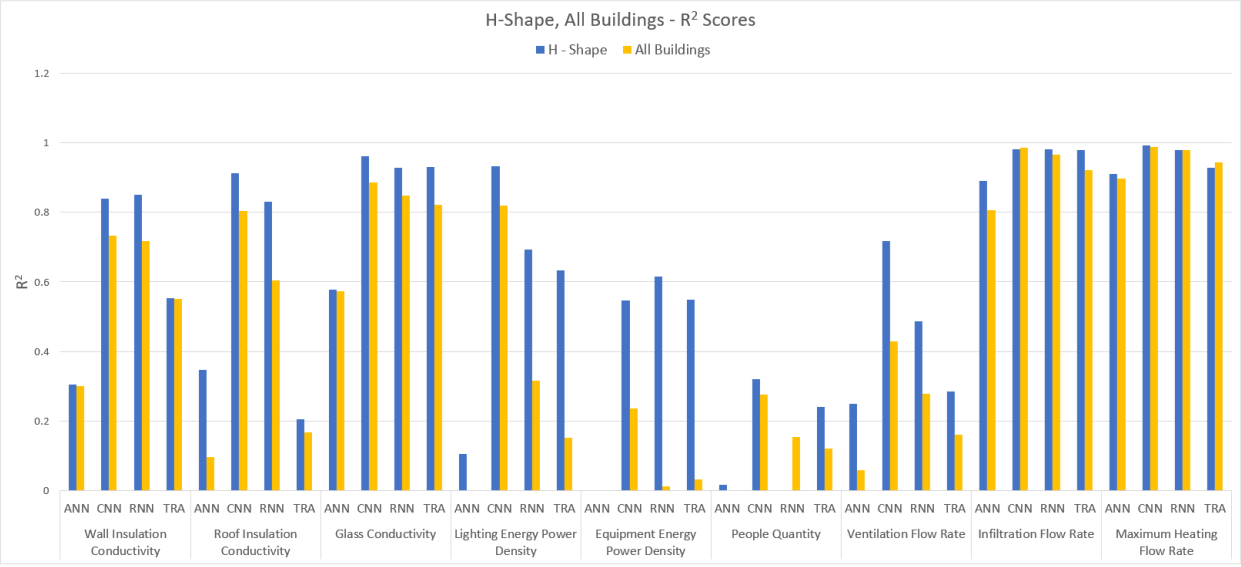


Figure 3.16 H-Shape, All Buildings - R^2 scores. *While prediction performance decreased when switched to training for all buildings, prediction performance with the higher performing parameters remained high.*

While weaker, the ISMs still retain the ability to make adequate and sometimes strong predictions for several parameters. Like the results provided in Figures 3.14 and 3.15, the Infiltration Flow Rate, Maximum Heating Flow Rate and Glass Conductivity remain suitable for the ISMs to predict. As only 2000 samples of each BEM model were used for training with all BEM models at once (compared to a full 5000 when trained individually), expanding the training dataset to include more samples of each model may demonstrate a performance benefit.

3.5 Conclusions

This study examined the generalizability of ISMs in their ability to predict BEM parameters from typical BEM output data. Each of the three BEM models from which training data was generated had scalable geometry requiring the ISMs to be able to decipher parameter values regardless of the exact shape of the building. BEM output data to be fed into the ISM was structured as a collection of time series data in the form of decay curves. Overall, the CNN preformed the strongest with the RNN and Transformer preforming similarly. The ANN, while a suitable benchmark, struggled to make out the influence of different parameters and therefore its performance was more limited to only being able to predict the values accurately for a small number of parameters. Additionally we examined the prediction performance of the ISMs when trained on a mixture of decay and rising curves, as well as various locations. It was found that the ISMs typically performed more strongly with a combination of decay and rising curves, however, this may be in large part due to the increase in the number of time series provided. Testing with variable locations, as well as training the ISMs on all BEM models at once, demonstrated a reduction in performance, but predictions with several parameters were still satisfactory for the CNN, RNN and Transformer.

In terms of future work, exploring modifications to hyperparameters and the ML model construction for each of the ISMs would potentially prove insightful. While an ideal hyperparameter and model construction combination for one BEM model may not be ideal for another, an examination of this in some form would be beneficial for comparisons between ISMs as well as demonstrate their resilience to further changes. Additional further examination into the

predictability of lesser predictable parameters would provide insight into what caused the prediction performance on them to notably fluctuate between BEM models in this study.

Overall, ISMs have been shown to provide a promising means of placing accurate predictions on BEM parameters even when the building geometry is varied. The prediction accuracy can vary considerably depending on the ISM model used in addition to the parameter being predicted. Prediction accuracy decreased when the ISMs were trained on data from multiple locations or on each core building geometry at once, however it still remained sufficient for several parameters.

Chapter 4 Conclusions and Future Work

With the growing age of the building stock nationwide and the substantial amount of green house gas emissions produced from buildings, ambitious targets are needed to reduce their environmental footprint. Building energy retrofits provide a means of improving the energy performance of an existing building. Compared to a complete teardown and rebuilding retrofits are often less costly, require less time to complete and result in less embodied carbon. To ensure that a retrofit is as effective as possible, understanding the characteristics of the building is essential.

Inverse surrogate modelling presents a new computational method of determining building characteristics. Surrogate modelling has emerged as an alternative to BEM, receiving great attention in recent years due to its near instant ability to predict BEM outputs. The development of ISMs is less widespread due to the more difficult task of determining inputs from outputs. This requires a careful selection of the ML model to be used, data quantity and data quality.

Chapter 2 explored the potential of an CNN ISM when tasked with predicting BEM parameters from time series data produced from a BEM model based on a single-family home. The quantity and quality of the data was adjusted to understand how much it would affect the prediction performance. Results showed that parameters related to the building envelope were more predictable than others. Parameters with weaker predictions were also more greatly affected when the data was adjusted.

Chapter 3 examined the generalizability of multiple ISMs when tasked to predict BEM parameters from varied geometric BEM models. Additionally, the ISMs were tasked to make accurate predictions when a BEM model had varied locations or were being trained on data from each BEM model at once. As with Chapter 2, building envelope parameters were the most predictable. Individual scaling on the BEM models still allowed for strong performance for the CNN, RNN and transformer, while performance expectedly decreased when trained for the multiple locations or a combination of all buildings.

In conclusion, this research demonstrated the potential of ISMs for the tasks of determining BEM parameters. ISMs were shown to be robust in handling different quantities and qualities of data while also being generalizable to multiple BEM models at a time.

Future work largely pertains to the expansion of this research for actual buildings. If it can be shown that ISMs can make accurate predictions of key characteristics, then building energy retrofits could be targeted more effectively. The studies in this thesis used synthetic data produced via BEM. Even with accurate geometry and loads, a significant number of differences remain, such as occupant behaviour, non-uniform building assemblies and unique HVAC systems. Understanding the effect of these differences and how to best incorporate them, such that the ISM can still train properly, presents a considerable challenge.

Additionally, when predicting BEM parameters, those that produced less of an effect on the internal temperatures were less consistently predictable. As demonstrated in Chapter 3, depending on the ISM used, as well as the BEM model from which the data was taken from, significantly altered the predictability of these parameters. Further investigation would help provide insight for why this occurs and what could be undertaken to ensure more consistent performance.

Chapter 5 References

[1] – Updated Framework Guidelines for Energy Efficiency Standards in Buildings., United Nations Economic and Social Council., Geneva, Switzerland, 2020. [Online]. Available: https://unece.org/sites/default/files/2020-12/ECE_ENERGY_GE.6_2020_4e.pdf

[2] – "Net-Zero Emissions by 2050" Canada.com https://www.canada.ca/en/services/environment/weather/climatechange/climate-plan/net-zero-emissions-2050.html (accessed October 7, 2023)

[3] – "2030 Emissions Reduction Plan – Sector-by-sector overview" Canada.com https://www.canada.ca/en/services/environment/weather/climatechange/climate-plan/climate-plan-overview/emissions-reduction-2030/sector-overview.html#sector2 (accessed October 7, 2023)

[4] – "Census Profile, 2021 Census of Population" statcan.gc.ca. https://www12.statcan.gc.ca/census-recensement/2021/dp-pd/prof/index.cfm?Lang=E (accessed October 7, 2023)

[5] - H. Mohammadpourkarbasi, B. Riddle, C. Liu, and S. Sharples, "Life cycle carbon assessment of decarbonising UK's hard-to-treat homes: A comparative study of conventional retrofit vs EnerPHit, heat pump first vs fabric first and ecological vs petrochemical retrofit approaches," *Energy and buildings*, vol. 296, p. 113353–, 2023, doi: 10.1016/j.enbuild.2023.113353.

[6] - D. Urge-Vorsatz, K. Petrichenko, M. Staniec, and J. Eom, "Energy use in buildings in a long-term perspective," *Current opinion in environmental sustainability*, vol. 5, no. 2, pp. 141–151, 2013, doi: 10.1016/j.cosust.2013.05.004.

[7] - H. Burak Gunay, D. Darwazeh, S. Shillinglaw, and I. Wilton, "Remote characterization of envelope performance through inverse modelling with building automation system data," *Energy and buildings*, vol. 240, p. 110893–, 2021, doi: 10.1016/j.enbuild.2021.110893.

[8] - G. Baasch, *Identification of thermal building properties using gray box and deep learning methods*. 2020.

[9] - H. Gao, C. Koch, and Y. Wu, "Building information modelling based building energy modelling: A review," *Applied energy*, vol. 238, pp. 320–343, 2019, doi: 10.1016/j.apenergy.2019.01.032.

[10] - A. Rasooli and L. Itard, "Automated in-situ determination of buildings' global thermo-physical characteristics and air change rates through inverse modelling of smart meter and air temperature data," *Energy and buildings*, vol. 229, p. 110484–, 2020, doi: 10.1016/j.enbuild.2020.110484.

[11] - A. Chong, Y. Gu, and H. Jia, "Calibrating building energy simulation models: A review of the basics to guide future work," *Energy and buildings*, vol. 253, p. 111533–, 2021, doi: 10.1016/j.enbuild.2021.111533.

[12] - P. Westermann and R. Evins, “Surrogate modelling for sustainable building design – A review,” *Energy and buildings*, vol. 198, pp. 170–186, 2019, doi: 10.1016/j.enbuild.2019.05.057.

[13] - S. Nagpal, C. Mueller, A. Aijazi, and C. F. Reinhart, “A methodology for auto-calibrating urban building energy models using surrogate modeling techniques,” *Journal of building performance simulation*, vol. 12, no. 1, pp. 1–16, 2019, doi: 10.1080/19401493.2018.1457722.

[14] - F. Herbinger, C. Vandenhof, and M. Kummert, “Building energy model calibration using a surrogate neural network,” *Energy and buildings*, vol. 289, p. 113057–, 2023, doi: 10.1016/j.enbuild.2023.113057.

[15] - F. Ascione, N. Bianco, C. De Stasio, G. M. Mauro, and G. P. Vanoli, “Artificial neural networks to predict energy performance and retrofit scenarios for any member of a building category: A novel approach,” *Energy (Oxford)*, vol. 118, pp. 999–1017, 2017, doi: 10.1016/j.energy.2016.10.126.

[16] - S. A. Sharif and A. Hammad, “Developing surrogate ANN for selecting near-optimal building energy renovation methods considering energy consumption, LCC and LCA,” *Journal of Building Engineering*, vol. 25, p. 100790–, 2019, doi: 10.1016/j.jobe.2019.100790.

[17] - E. Asadi, M. G. da Silva, C. H. Antunes, L. Dias, and L. Glicksman, “Multi-objective optimization for building retrofit: A model using genetic algorithm and artificial neural network and an application,” *Energy and buildings*, vol. 81, pp. 444–456, 2014, doi: 10.1016/j.enbuild.2014.06.009.

[18] - F. Shadram, S. Bhattacharjee, S. Lidelöw, J. Mukkavaara, and T. Olofsson, “Exploring the trade-off in life cycle energy of building retrofit through optimization,” *Applied energy*, vol. 269, p. 115083–, 2020, doi: 10.1016/j.apenergy.2020.115083.

[19] - Y.-D. Ko and C.-S. Park, “Parameter estimation of unknown properties using transfer learning from virtual to existing buildings,” *Journal of building performance simulation*, vol. 14, no. 5, pp. 503–514, 2021, doi: 10.1080/19401493.2021.1972159.

[20] - T. Hong and S. H. Lee, “Integrating physics-based models with sensor data: An inverse modeling approach,” *Building and environment*, vol. 154, no. C, pp. 23–31, 2019, doi: 10.1016/j.buildenv.2019.03.006.

[21] – “eSim2022” Carleton.ca. https://carleton.ca/esim22/en_homepage/ (accessed October 7, 2023)

[22] – “Building Simulation 2023” bs2023.org. https://bs2023.org/ (accessed October 7, 2023)

[23] – “EnergyPlus” energyplus.net. https://energyplus.net/ (accessed October 7, 2023)

[24] – “Group – Zone Forced Air Units” bigladdersoftware.com https://bigladdersoftware.com/epx/docs/8-0/input-output-reference/page-032.html#zonehvacidealloadsairsystem (accessed October 7, 2023)

[25] – “Publication” bs2023.org. https://bs2023.org/publication (accessed October 7, 2023)

[26] - I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*. Cambridge, Massachusetts: The MIT Press, 2016.

[27] - M. D. Shields and J. Zhang, "The generalization of Latin hypercube sampling," *Reliability engineering & system safety*, vol. 148, pp. 96–108, 2016, doi: 10.1016/j.ress.2015.12.002.

[28] – "6.3. Preprocessing data." scikit-learn.org https://scikit-learn.org/stable/modules/preprocessing.html

[29] – U. Michelucci, *Advanced Applied Deep Learning Convolutional Neural Networks and Object Detection*, 1st ed. 2019. Berkeley, CA: Apress, 2019. doi: 10.1007/978-1-4842-4976-5.

[30] - S. Kiranyaz, O. Avci, O. Abdeljaber, T. Ince, M. Gabbouj, and D. J. Inman, "1D convolutional neural networks and applications: A survey," *Mechanical systems and signal processing*, vol. 151, p. 107398–, 2021, doi: 10.1016/j.ymssp.2020.107398.

[31] – P. Westermann, M. Welzel, and R. Evins, "Using a deep temporal convolutional network as a building energy surrogate model that spans multiple climate zones," *Applied energy*, vol. 278, p. 115563–, 2020, doi: 10.1016/j.apenergy.2020.115563.

[32] - D. Chakraborty and H. Elzarka, "Performance testing of energy models: are we using the right statistical metrics?," *Journal of building performance simulation*, vol. 11, no. 4, pp. 433–448, 2018, doi: 10.1080/19401493.2017.1387607.

[33] - T. Østergård, R. L. Jensen, and S. E. Maagaard, "A comparison of six metamodeling techniques applied to building performance simulations," *Applied energy*, vol. 211, pp. 89–103, 2018, doi: 10.1016/j.apenergy.2017.10.102.

[34] - T. Chai and R. R. Draxler, "Root mean square error (RMSE) or mean absolute error (MAE)? – Arguments against avoiding RMSE in the literature," *Geoscientific model development*, vol. 7, no. 3, pp. 1247–1250, 2014, doi: 10.5194/gmd-7-1247-2014.

[35] – ASHRAE, "Guideline 14-2014, Measurement of Energy and Demand Savings," Atlanta, GA, USA, Dec. 2014.

[36] – G. Chaudhary, J. New, J. Sanyal, P. Im, Z. O'Neill, and V. Garg, "Evaluation of 'Autotune' calibration against manual calibration of building energy models," *Applied energy*, vol. 182, no. C, pp. 115–134, 2016, doi: 10.1016/j.apenergy.2016.08.073.

[37] – H. Wang and Z. (John) Zhai, "Advances in building simulation and computational techniques: A review between 1987 and 2014," *Energy and buildings*, vol. 128, pp. 319–335, 2016, doi: 10.1016/j.enbuild.2016.06.080.

[38] – "National Housing Strategy." Cmhc-schl.gc.ca.

https://www.cmhc-schl.gc.ca/nhs/guidepage-strategy (accessed Feb. 12, 2024).

[39] – *2022 Global Status Report for Buildings and Construction*., United Nations Environment Programme. Geneva, Switzerland, 2022

[40] – “VE Virtual Environment.” IESVE.com. https://www.iesve.com/software/virtual-environment (accessed Feb. 12, 2024)

[41] – “Design Builder.” designbuilder.co.uk. https://designbuilder.co.uk/ (accessed Feb 12. 2024)

[42] - H. Gao, C. Koch, and Y. Wu, “Building information modelling based building energy modelling: A review,” *Applied energy*, vol. 238, pp. 320–343, 2019, doi: 10.1016/j.apenergy.2019.01.032.

[43] - E. Fabrizio and V. Monetti, “Methodologies and Advancements in the Calibration of Building Energy Models,” *Energies*, vol. 8, no. 4, pp. 2548–2574, 2015, doi: 10.3390/en8042548.

[44] – “PathFinder.” Buildingpathfinder.com. https://www.buildingpathfinder.com/ (accessed Feb 12.2024)

[45] - A. Chong, Y. Gu, and H. Jia, “Calibrating building energy simulation models: A review of the basics to guide future work,” *Energy and buildings*, vol. 253, pp. 111533-, 2021, doi: 10.1016/j.enbuild.2021.111533.

[46] - D. Hou and R. Evins, “A protocol for developing and evaluating neural network-based surrogate models and its application to building energy prediction,” *Renewable & sustainable energy reviews*, vol. 193, pp. 114283-, 2024, doi: 10.1016/j.rser.2024.114283.

[47] – C. Janiesch, P. Zschech, and K. Heinrich, “Machine learning and deep learning,” *Electronic markets*, vol. 31, no. 3, pp. 685–695, 2021, doi: 10.1007/s12525-021-00475-2.

[48] – P. W. Tien, S. Wei, J. Darkwa, C. Wood, and J. K. Calautit, “Machine Learning and Deep Learning Methods for Enhancing Building Energy Efficiency and Indoor Environmental Quality – A Review,” *Energy and AI*, vol. 10, pp. 100198-, 2022, doi: 10.1016/j.egyai.2022.100198.

[49] - Daniel. Graupe, *Principles of artificial neural networks*, 3rd ed. Singapore ; World Scientific, 2013.

[50] - A. Géron, *Hands-on machine learning with Scikit-learn, Keras, and TensorFlow : concepts, tools, and techniques to build intelligent systems*, 2nd ed. Sebastopol: O’Reilly, 2019.

[51] – Razak Olu-Ajayi, Hafiz Alaka, Ismail Sulaimon, Funlade T. Sunmola, and Saheed O. Ajayi, “Building energy consumption prediction for residential buildings using deep learning and other machine learning techniques,” *Journal of Building Engineering*, vol. 45, pp. 103406-, 2022, doi: 10.1016/j.jobe.2021.103406.

[52] – G. Suryanarayana, J. Lago, D. Geysen, P. Aleksiejuk, and C. Johansson, “Thermal load forecasting in district heating networks using deep learning and advanced feature selection methods,” Energy (Oxford), vol. 157, pp. 141–149, 2018, doi: 10.1016/j.energy.2018.05.111.

[53] - F. M. Salem, *Recurrent Neural Networks: From Simple to Gated Architectures*, 1st ed. Cham: Springer International Publishing AG, 2022. doi: 10.1007/978-3-030-89929-5.

[54] - C. Fan, J. Wang, W. Gang, and S. Li, "Assessment of deep recurrent neural network-based strategies for short-term building energy predictions," *Applied energy*, vol. 236, pp. 700–710, 2019, doi: 10.1016/j.apenergy.2018.12.004.

[55] - S. Jung, J. Moon, S. Park, and E. Hwang, "An Attention-Based Multilayer GRU Model for Multistep-Ahead Short-Term Load Forecasting," *Sensors (Basel, Switzerland)*, vol. 21, no. 5, pp. 1639-, 2021, doi: 10.3390/s21051639.

[56] - J. Jang, J. Han, and S.-B. Leigh, "Prediction of heating energy consumption with operation pattern variables for non-residential buildings using LSTM networks," *Energy and buildings*, vol. 255, pp. 111647-, 2022, doi: 10.1016/j.enbuild.2021.111647.

[57] – N. Somu, G. Raman M R, and K. Ramamritham, "A deep learning framework for building energy consumption forecast," *Renewable & sustainable energy reviews*, vol. 137, pp. 110591-, 2021, doi: 10.1016/j.rser.2020.110591.

[58] – U. Kamath, *Transformers for machine learning : a deep dive*, First edition. Boca Raton, Florida: CRC Press, 2022. doi: 10.1201/9781003170082.

[59] ANSI/ASHRAE Addendum a to ANSI/ASHRAE Standard 169-2020 – Climatic Data for Building Design Standards, 169-2020, ASHRAE, Atlanta, GA, USA, Oct. 2021. [Online]. Available: https://www.ashrae.org/file%20library/technical%20resources/standards%20and%20guidelines/standards%20addenda/169_2020_a_20211029.pdf

[60] – Vaswani, Ashish, et al. "Attention is all you need." Advances in neural information processing systems., June. 2017, doi: 10.48550/arXiv.1706.03762

[61] D. S. K. Karunasingha, "Root mean square error or mean absolute error? Use their ratio as well," *Information sciences*, vol. 585, pp. 609–629, 2022, doi: 10.1016/j.ins.2021.11.036.

[62] – T. O. Hodson, "Root-mean-square error (RMSE) or mean absolute error (MAE): when to use them or not," *Geoscientific Model Development*, vol. 15, no. 14, pp. 5481–5487, 2022, doi: 10.5194/gmd-15-5481-2022.

[63] A. de Myttenaere, B. Golden, B. Le Grand, and F. Rossi, "Mean Absolute Percentage Error for regression models," *Neurocomputing (Amsterdam)*, vol. 192, pp. 38–48, 2016, doi: 10.1016/j.neucom.2015.12.114.

Chapter 6 Appendix
6.1 Appendix A Error Metric Tables

Table A.1 C-Shape Error Metric Scores

C-Shape	R^2				MAPE				RMSPE			
Scenario	ANN	CNN	RNN	TRA	ANN	CNN	RNN	TRA	ANN	CNN	RNN	TRA
Wall Insulation Conductivity	0.39	0.90	0.84	0.78	43.2	14.4	19.2	20.6	68.2	22.2	29.8	29.6
Roof Insulation Conductivity	0.28	0.75	0.74	0.44	53.4	31.2	26.3	46.6	92.2	54.9	46.1	80.1
Glass Conductivity	0.66	0.94	0.92	0.89	24.6	9.4	9.7	12.4	35.6	12.3	12.9	16.7
Lighting Energy Power Density	0.00	0.91	0.82	0.79	10.3	2.5	3.8	4.1	12.1	3.4	5.0	5.3
Equipment Energy Power Density	-0.01	0.31	0.11	0.04	10.1	7.7	9.5	9.8	11.9	9.6	11.3	11.8
People Quantity	0.05	0.35	0.29	0.28	27.6	23.7	23.2	23.7	35.5	32.2	30.8	32.3
Ventilation Flow Rate	0.07	0.39	0.43	0.29	16.8	12.2	12.5	14.4	20.5	15.0	15.4	18.3
Infiltration Flow Rate	0.88	0.99	0.98	0.96	19.3	4.6	7.6	9.5	32.7	7.3	12.3	13.6
Maximum Heating Flow Rate	0.92	0.99	0.98	0.98	13.4	5.1	7.0	8.3	19.7	9.0	10.4	12.2
Scale - X	0.27	0.96	0.89	0.87	>100	16.4	27.6	36.0	>100	39.7	94.7	>100
Scale - Y	0.13	0.91	0.71	0.73	>100	24.5	54.9	38.8	>100	52.9	>100	74.5
Scale - Z	0.71	0.95	0.95	0.91	68.8	23.2	26.6	28.9	>100	66.5	91.0	67.9
Range	0.00		1.00		0.0		70.0		0.0		50.0	

Table A.2 Octagonal-Shape Error Metric Scores

Octagonal-Shape	R^2				MAPE				RMSPE			
Scenario	ANN	CNN	RNN	TRA	ANN	CNN	RNN	TRA	ANN	CNN	RNN	TRA
Wall Insulation Conductivity	0.59	0.96	0.90	0.87	35.8	10.4	13.8	17.3	60.5	15.6	19.9	27.5
Roof Insulation Conductivity	0.10	0.77	0.80	0.87	65.9	23.9	20.4	18.5	>100	31.3	28.7	27.5
Glass Conductivity	0.68	0.97	0.92	0.88	23.0	6.7	9.8	11.8	31.0	8.9	12.7	15.2
Lighting Energy Power Density	-0.01	0.75	0.33	0.31	9.9	4.5	7.3	7.7	11.7	5.7	9.3	9.6
Equipment Energy Power Density	0.01	0.42	0.00	0.02	9.7	6.8	9.4	9.8	11.6	8.5	11.8	12.0
People Quantity	0.01	0.47	0.31	0.21	28.2	19.6	20.7	24.3	37.3	27.5	28.5	32.9
Ventilation Flow Rate	0.06	0.43	0.26	0.23	16.9	12.0	14.1	14.9	20.9	15.5	17.7	18.5
Infiltration Flow Rate	0.81	0.98	0.97	0.93	21.3	7.6	8.9	13.1	32.4	12.4	13.8	19.0
Maximum Heating Flow Rate	0.91	0.99	0.98	0.98	15.3	4.9	6.8	7.1	24.0	8.1	11.2	11.3
Scale - X	0.41	0.80	0.78	0.76	>100	42.5	51.4	51.1	>100	>100	>100	>100
Scale - Y	0.49	0.89	0.86	0.76	98.6	37.2	39.5	41.3	>100	95.6	>100	>100
Scale - Z	0.68	0.95	0.91	0.87	61.2	19.0	27.7	27.7	>100	57.2	>100	59.5
Range	0.00		1.00		0.0		70.0		0.0		50.0	

Table A.3 Decay and Rise Error Metric Scores

Decay and Rise	R^2				MAPE				RMSPE			
Scenario	ANN	CNN	RNN	TRA	ANN	CNN	RNN	TRA	ANN	CNN	RNN	TRA
Wall Insulation Conductivity	0.48	0.88	0.87	0.90	39.2	18.0	17.2	16.9	63.6	30.0	27.4	26.0
Roof Insulation Conductivity	0.74	0.83	0.88	0.80	29.6	24.9	18.1	22.6	50.8	43.9	33.6	29.7
Glass Conductivity	0.77	0.95	0.94	0.97	21.3	8.7	9.2	6.0	30.4	12.0	12.4	7.9
Lighting Energy Power Density	0.34	0.81	0.76	0.89	8.1	4.0	4.3	3.0	10.3	5.0	5.9	4.0
Equipment Energy Power Density	0.11	0.74	0.58	0.69	9.1	4.6	6.0	5.9	11.2	6.0	7.5	6.7
People Quantity	0.07	0.30	0.08	0.34	27.0	22.7	27.2	23.7	34.1	28.3	36.7	31.5
Ventilation Flow Rate	0.35	0.68	0.59	0.59	13.4	9.1	9.7	9.8	17.0	11.9	12.8	12.5
Infiltration Flow Rate	0.93	0.99	0.98	0.98	13.3	5.1	6.3	7.3	21.0	7.9	9.7	11.6
Maximum Heating Flow Rate	0.94	0.99	0.99	0.99	11.9	4.4	5.9	5.5	17.1	5.9	8.7	7.9
Scale - X	0.72	0.88	0.92	0.93	65.5	43.6	27.0	26.1	>100	>100	80.7	57.9
Scale - Y	0.11	0.70	0.57	0.88	>100	46.9	68.0	27.2	>100	>100	>100	58.1
Scale - Z	0.85	0.95	0.96	0.98	54.0	22.4	17.4	13.3	>100	51.6	37.7	34.7
Range	0.00		1.00		0.0		70.0		0.0		50.0	

Table A.4 Variable Location Metric Scores

Variable Location	R^2				MAPE				RMSPE			
Scenario	ANN	CNN	RNN	TRA	ANN	CNN	RNN	TRA	ANN	CNN	RNN	TRA
Wall Insulation Conductivity	0.06	0.61	0.44	0.61	63.9	28.5	44.7	29.9	>100	41.5	78.2	47.0
Roof Insulation Conductivity	0.30	0.83	0.53	0.73	51.0	23.4	40.2	29.4	85.5	41.4	69.3	53.4
Glass Conductivity	0.56	0.89	0.83	0.88	29.4	12.7	15.4	12.1	43.4	17.8	22.4	15.5
Lighting Energy Power Density	0.05	0.65	0.15	0.35	9.7	5.9	8.9	7.7	11.6	7.3	10.8	9.7
Equipment Energy Power Density	-0.06	0.22	0.06	0.35	10.2	8.0	9.5	7.7	12.2	9.7	11.8	9.6
People Quantity	-0.02	0.21	0.01	0.00	28.4	22.8	25.5	26.4	36.6	29.5	32.2	34.8
Ventilation Flow Rate	0.05	0.44	0.19	0.27	16.6	11.4	15.3	14.1	20.5	14.7	19.5	18.2
Infiltration Flow Rate	0.82	0.95	0.94	0.97	24.9	13.2	12.5	9.3	41.7	20.2	20.2	14.1
Maximum Heating Flow Rate	0.92	0.95	0.97	0.98	16.7	11.5	9.0	7.1	28.6	15.2	15.3	11.0
Scale - X	0.25	0.87	0.68	0.94	>100	38.4	61.2	22.4	>100	98.2	>100	50.6
Scale - Y	-0.01	0.53	0.17	0.53	>100	91.3	>100	91.1	>100	>100	>100	>100
Scale - Z	0.64	0.94	0.88	0.96	78.4	22.8	35.7	17.3	>100	54.4	97.2	43.4
Latitude	0.95	0.99	0.98	0.98	4.7	1.9	2.6	3.1	6.0	2.6	3.4	3.9
Longitude	0.94	0.97	0.97	0.94	3.8	2.9	2.4	3.7	5.1	3.4	3.1	4.6
Range	0.00		1.00		0.0		70.0		0.0		50.0	

Table A.5 All Buildings Metric Scores

All Buildings	R^2				MAPE				RMSPE			
Scenario	ANN	CNN	RNN	TRA	ANN	CNN	RNN	TRA	ANN	CNN	RNN	TRA
Wall Insulation Conductivity	0.30	0.73	0.72	0.55	50.8	30.3	28.4	35.0	82.8	46.6	44.8	54.1
Roof Insulation Conductivity	0.10	0.80	0.60	0.17	55.0	21.8	31.5	48.2	89.3	37.2	53.8	76.4
Glass Conductivity	0.57	0.89	0.85	0.82	27.4	12.9	14.1	15.1	38.3	16.4	18.9	20.0
Lighting Energy Power Density	-0.01	0.82	0.32	0.15	10.2	3.6	7.8	8.7	12.0	5.2	9.8	10.8
Equipment Energy Power Density	-0.01	0.24	0.01	0.03	9.8	7.7	9.5	9.3	11.7	9.6	11.6	11.4
People Quantity	0.00	0.28	0.16	0.12	28.9	21.2	25.9	26.5	37.9	27.6	34.6	36.2
Ventilation Flow Rate	0.06	0.43	0.28	0.16	17.0	12.4	13.9	15.6	21.0	16.1	17.5	19.9
Infiltration Flow Rate	0.81	0.99	0.97	0.92	25.0	6.3	9.3	13.8	40.5	10.7	14.9	21.2
Maximum Heating Flow Rate	0.90	0.99	0.98	0.94	16.4	5.4	6.5	11.7	27.6	8.8	9.9	16.6
Scale - X	0.25	0.90	0.78	0.74	>100	36.2	52.9	58.7	>100	>100	>100	>100
Scale - Y	0.18	0.75	0.47	0.45	>100	67.4	>100	>100	>100	>100	>100	>100
Scale - Z	0.63	0.95	0.89	0.86	97.8	26.0	41.7	41.8	>100	80.1	>100	>100
Range	0.00		1.00		0.0		70.0		0.0		50.0	

6.2 Appendix B Using a Convolutional Neural Network to Determine the Thermal Characteristics of a Building

Using a Convolutional Neural Network to Determine the Thermal Characteristics of a Building

Liam Jowett-Lockwood[1,2], Ralph Evins[1,2]
[1]Energy in Cities group, Department of Civil Engineering, University of Victoria, BC, Canada
[2]Institute for Integrated Energy Systems, University of Victoria, BC, Canada

Abstract

Surrogate models are machine learning models that are trained using detailed simulation input and output data and can practically provide instant results. They use a forward implementation strategy and normally do not use building simulation outputs to identify inputs. This paper introduces inverse surrogate modeling of a building by using a convolutional neural network that uses temperature data to estimate the building characteristics, such as the wall insulation conductivity and infiltration flow rate. The training data resembles thousands of 10-minute interval temperature time series along with their respective building parameters. The first proof of principle uses synthetic data from a building energy model to train and test the neural network. This will later predict actual building parameters using real temperature data from an existing building. Findings demonstrate that combining temperature data with a convolutional neural network could serve as a cost-effective method of building parameter identification.

Introduction

Many cities in the Canadian province of British Columbia contain a significant number of homes built close to or over half a century ago. The 2016 Canadian census states that based on a sample of 25% of occupied private dwellings province-wide, over 43% of homes were built prior to 1981 and over 14% were built prior to 1961 (Statistics Canada (2016)). This number is greater for older cities, such as the city of Victoria which has percentages of 56% and 24% for homes built prior to 1981 and 1961 respectively. Due to the age of these homes, information regarding their parameters (e.g., wall insulation U-value, electric equipment energy) may not be recorded and therefore not readily available for the current owner or potential buyer. Furthermore, some parameters, such as the infiltration flow rate, require tests that are expensive or are intrusive for building occupants (Edwards et al. (2017)). Lack of knowledge on key parameters makes it difficult to prioritize retrofit measures and estimate their benefits.

Buildings are a significant contributor to climate change with a contribution of one-third of global GHG emissions (Bamdad et al. (2021)). A significant portion of these emissions is related to meeting the operational energy demand composed of heating, cooling and electricity use. Effective retrofit strategies do exist, however, that can substantially influence the required energy demand. Due to variability in home construction, climate and deterioration, consideration should be applied to determine whether retrofit strategies are an effective means of reducing emissions and providing long-term cost savings. Not applying sufficient consideration may result in some unfortunate scenarios, where completed retrofits produce little benefit and the required extent of retrofits needed may be too uneconomical to justify (Urge-Vorsatz et al. (2013)). Even if some retrofit strategies are not viable, homes may have sizeable energy usage in other areas, including high equipment and lighting energy usage. Determining whether or not performing retrofits are suitable or if other options should be taken can be challenging, especially if information is unknown.

One approach to determine a building's parameter information would be to use a building energy simulation model to replicate the building energy performance. However, modifying parameter values and producing multiple simulation runs to match the energy performance to that of the actual building has two immediate issues. Firstly, there is the case where the energy performance is matched by multiple different sets of parameter values. This would suggest that some parameters should be found via other means to remove or at least reduce this occurrence. Using outputs of a different form may also help alleviate this issue (e.g., using a time series of energy usage instead of a single aggregate value). Secondly, completing multiple building simulation model runs is computationally intensive and requires significant time. For this point, a potential alternative would be the use of a Surrogate Model (SM).

A SM is a machine learning model designed to repli-

cate the output(s) produced from a high-fidelity model (Westermann and Evins (2019)). For a building energy simulation model, the SM is trained on the model's inputs and outputs, such that the model can accurately predict outputs from unseen inputs. For synthetic data, the process often involves the development of a building simulation model to produce extensive training data. While the development of a building simulation model for the purposes of training a SM may seem to make use of the SM nonsensical, ideally a trained SM can be applied to other projects/geometries as well. Research has shown that when developed effectively, a SM can be deployed to achieve strong prediction performance where multiple building designs are trained for and the location is held constant (Lawrence et al. (2021)).

For buildings, SMs have several different uses including parameter sensitivity analysis, early design exploration, and optimization (Westermann and Evins (2019)). Furthermore, the broad field of surrogate modelling is composed of many different types of machine learning models including linear regression, Gaussian process and Artificial Neural Network (ANN) models. In particular, a Convolutional Neural Network (CNN) is an ANN with the unique implementation of convolutional kernels. This allows for a degree of invariance in the model as kernel weights are reused at each section of input. The usage of convolutional kernels allows CNNs to more easily handle potentially extensive time series inputs, such as building energy usage or temperature recordings, in comparison to other machine learning models (Westermann et al. (2020)). This would thereby suggest it as a strong candidate as a SM for projects involving inputs of this scale.

SMs for buildings typically use building specific parameters as model inputs and energy use or simulated results as outputs. Creating an inverse building SM would involve a machine learning model that uses typical building simulation outputs as inputs with the intention of predicting building parameters. With the effective ability of CNNs to accurately learn from extensive time-series inputs, there is a potential to treat building time series information, such as hourly energy usage or zone temperature, as an inverse SM input. For existing buildings, time series outputs may be known from data from recent years or easily collected sensor data. These can be used to determine unknown building parameters through the use of an Inverse Surrogate Model (ISM).

This paper presents the development and evaluation of a building ISM to facilitate the determination of specific building parameters. The ISM is based on a deep CNN because of the strong performance of CNNs with time series inputs. Trained on multi-zone building temperature data, this CNN ISM attempts to identify building parameters that may influence the temperature of different rooms over time. This study currently only examines the effectiveness of the model on synthetic data, however it is intended to eventually extend the research to include real data. This would resemble the model being training on temperature outputs from a building simulation model designed on a real structure and will then predict the actual building parameters of the same structure using sensor data.

This paper is organized as follows: Section 2 describes the main elements of the overarching methodology to formulate trained inverse building surrogate models. Section 3 describes and discusses the findings and adjustments made to the methods in analyzing results. Section 4 discuss the planned future work of incorporating real data. Section 5 concludes the paper.

Methodology

The process demonstrating the overall workflow is shown in Figure 1. The two most significant portions of this process was the development of the synthetic data generation and ISM. Data pre-processing is applied to the data prior to training to improve performance. Various error metrics are used to evaluate synthetic data, however they will be omitted from eventual real data evaluation due to the lack of reference data.

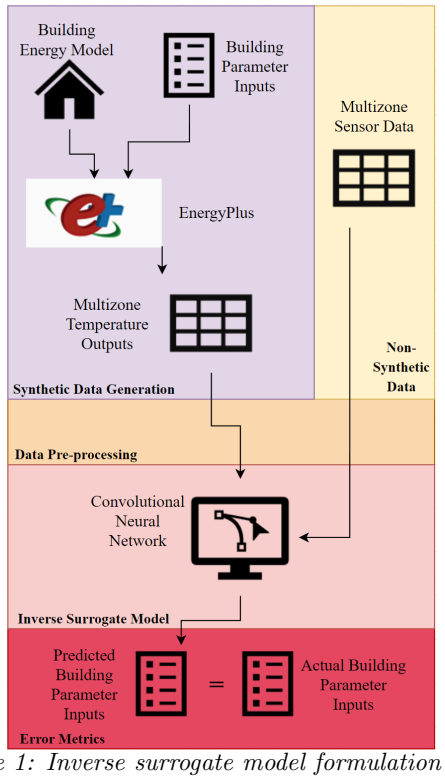


Figure 1: Inverse surrogate model formulation workflow to predict building parameter inputs.

For predicting synthetic data, the process can almost be considered circular, as varied parameter inputs are initially fed into the building simulation model and the final output of the ISM are predictions of the original inputs.

Synthetic Data Generation

In order for the trained ISM to be best suited for handling real data, the building simulation model was designed to reflect a specific existing residential building in Victoria, British Columbia. The building simulation software EnergyPlus was used for the simulation modeling. Within the simulation model, each of the 14 rooms of the home are treated as separate thermal zones from which temperature values are produced. Out of these rooms, 7 are present on the first floor, 5 on the second floor and 1 each for the basement and attic. Some of these rooms include hall and entry ways. For the privacy of the building owner and occupants, an exact floor plan and room layout have been omitted here. As some rooms are not fully separated from others (e.g. lack of a door), full interior partitions/walls are placed in the software to separate them.

In regards to HVAC, the EnergyPlus Ideal Air Loads HVAC system object was used in place of modelling mechanical systems. This object is typically used when the designer wishes to omit the implementation of a complicated HVAC system in the model while still wanting to examine building performance. The object resembles an ideal HVAC system that applies or removes heat and moisture at 100% efficiency to provide a supply air stream (USDOE (2022)). The unit is not connected to a central air system, rather each system object supplies air to a zone in necessary quantities to heat or cool in order to reach specified setpoints.

To produce training data for the ISM, various parameters were varied for each simulation run, see Table 1. The parameters and ranges were chosen based on previous research (Westermann et al. (2020)), their importance for a home retrofit and potential variability among households. As the considered building is an old construction (built in 1930s), values for many of the actual parameters are unknown. Engineering judgement was applied to develop assumptions on the properties of the constructions, however the extent of these assumptions is a likely contributor to a poor representation of the actual building. Randomized parameter value selection was conducted by using Latin Hypercube Sampling (LHS), which is a strong and widely used random sampling method because of its ability to greatly lower variance amongst different applications (Shields and Zhang (2015)).

Notably, both the Zone Ventilation Flow Rate and Zone Infiltration Flow Rate refer to the airflow in and out of the home. EnergyPlus also provides methods of transferring airflow around the home via mixing objects or its AirflowNetwork model, however these methods were not included in the model due to added complexity and how flow amounts would very significantly between different zones in the actual home because of opening sizes. Whereas the Zone Ventilation Flow Rate is defined as the purposeful flow of outside air to a thermal zone to provide cooling, the Zone Infiltration Flow Rate is the airflow into the zone through unintended openings, such as cracks around windows (USDOE (2022)). A significant difference between the two is that, for the Zone Infiltration Flow Rate, the flow per exterior surface area is what is specifically varied whereas the flow rate per person is varied for the Zone Ventilation Flow Rate. Calculations do not take into account location of envelope leakage. Furthermore, EnergyPlus does provide the option to use various formulas/models for airflow for both parameters, which would include the influence of additional factors, such as wind speed. However, for simplicity and the unknowns of the building itself, these have been omitted

Table 1: List of the varied building parameters for ISM training.

Parameter	Units	Range
North Axis	°	0-360
Wall Insulation Conductivity	W/mK	0.01-0.1
Roof Insulation Conductivity	W/mK	0.01-0.1
Foundation Concrete Thickness	m	0.05-0.2
Window Conductivity	W/mK	0.01-0.05
Lighting Energy	W/m^2	6-10
Electric Equipment Energy	W/m^2	5.5-9.5
Personnel	$People/m^2$	0.015-0.035
Zone Ventilation Flow Rate	L/sPerson	5-10
Zone Infiltration Flow Rate	L/m^2s	0.1-0.5

It is also noteworthy that, for some of the material parameters, conductivity or thickness were selected instead of U-value, so that the variation could be more easily implemented into the simulation model. However, the variation in U-value can still be obtained if need be.

In total 4000 simulation runs were completed, taking roughly a minute for each. Temperature values for each of the 14 zones were extracted with a Python script and organized in a CSV file resulting in 4000 separate data files. Temperature values are recorded in 10-minute intervals for a year which resulted in a total of 52560 values for each zone. Additionally included in the files was a column of binary values for which values of 1 indicated that heating was present and 0 indicating otherwise. For the purposes of training, this additional column is treated as another zone making the total 15 zones.

As each data file constitutes one simulation run, the decision to use only 4000 simulations was based on the sheer size of each file (15x52560 values). Out of the total 4000 files, the training, validation and testing sets were composed of 2560, 640 and 800 files respectively. The determination of these quantities was done by using a train-test split of 0.2 and a train-validation split of 0.2 as well. While the CNN may perform stronger with a larger amount of simulation data to train on, lowering the amount of data saves time and computation effort as less simulation runs are completed, and less data is used for ISM training.

While 52560 values were produced to encompass the entire year, training and testing occurred on 30-day intervals to determine if different time periods were impactful for results, hence only a portion of each file was used. The months chosen were March, June, September and December. The decision of choosing which months to use was based on the distance apart from each other and the time frame of the preliminary sensor data gathering to be eventually used for parameter prediction of the real structure.

Data Pre-processing

Prior to training the SM, pre-processing was applied to the data. This involved standardizing the dataset to resemble standard normally distributed data. Standardization re-scales the feature values into a more suitable range, such that the influence of differences between the values of data is less significant (P.Angelov and Gu (2019)). For this study Scikit-Learn's StandardScaler (Pedregosa et al. (2011)) functions were used to compute preprocessing for both the training and testing data. This involved centering the data by removing the mean value from each feature and scaling such that mean and variance values of 0 and 1 are achieved. To conduct the preprocessing it was required to load the entire data sets in memory. For the outputs (varied parameter values), these were simply loaded in as a single CSV file. For the inputs however, as these spanned multiple data files (one for each parameter variation) composed of multiple time series, the process involved loading each file and then modifying the two-dimensional matrix into a one-dimensional array as shown in Figure 2.

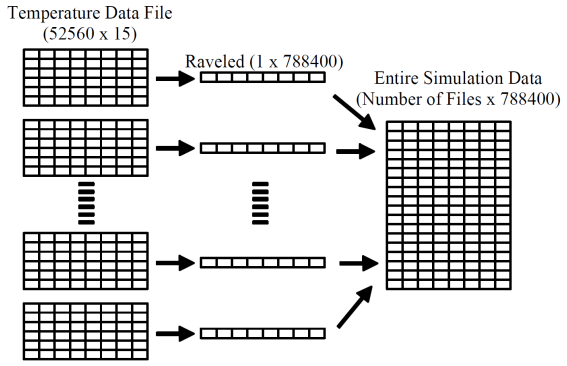


Figure 2: Pre-processing methodology for ISM input data where multiple files are loaded into memory as one.

It should be noted that the mean and standard deviation used for centering and scaling are calculated only from the samples in the training (and validation) set. The same mean and standard deviation values are then used on standardizing the testing set. Including the non-standardized testing set in the calculation of the mean and standard deviation is a form of data leakage and is a common mistake in machine learning, which can lead to false optimistic results.

Convolutional Neural Network Structure

The CNN structure is heavily based on the ResNet model architecture deployed by Westermann et al. (2020). The reader is referred to their research for a comprehensive breakdown of the various design decisions and underlying model features. The overview of the model structure is presented in Figure 3.

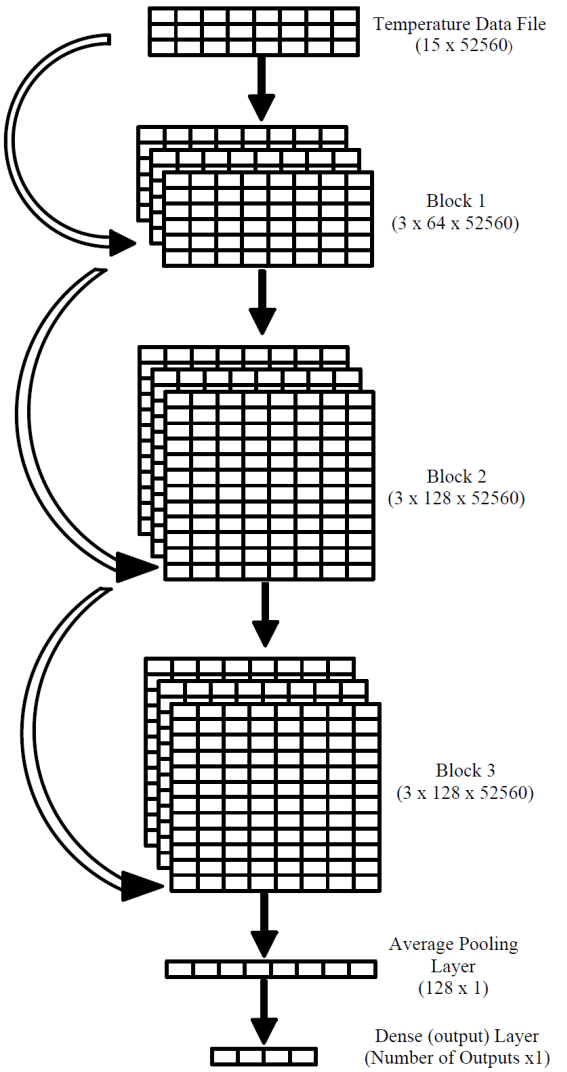


Figure 3: Convolutional neural network structure where straight arrows indicate a direct connection to subsequent blocks and curved arrows for residual shortcuts.

Each of the three blocks is composed of a one-dimensional convolutional, batch normalization and activation layer. The activation layer computes a Rectified Linear Unit (ReLU) activation. A residual shortcut is also included, thereby connecting the last layer of the previous block with the following block, in an effort remove the occurrence of the vanishing gradient. The main difference between the model developed by Westermann et al. (2020) is the removal of the inclusion of additional features after the final

dense layer. In their research, the ANN is continued with a concatenation of additional features to the dense layer which is followed by an additional dense layer and eventual output. This is unnecessary for this project as no features are trained for outside of the multiple inputted time series.

Error Metrics

The purpose of error metrics is to determine the effectiveness of the CNN in being able to predict parameters. After being trained, the ISM will predict building parameter values from the temperature time series data for each simulation run. The error metrics themselves are equations whose values serve as indicators of performance. The equations accept both the predicted building parameter values and actual as inputs. Assuming the building energy model is an accurate representation of the actual building, parameters that the ISM can predict with high accuracy with synthetic data would ideally achieve higher prediction accuracy than those with poorer performance when real data is used.

Various error metrics were used in this study. These match some of the metrics previously used by Westermann et al. (2020). The metrics consisted of the Coefficient of Determination (R^2), Mean Absolute Percentage Error (MAPE) and the Root Mean Squared Percentage Error (RMSPE), which are defined below:

$$R^2(y, \dot{y}) = 1 - \frac{\sum_{i=1}^{n}(y_i - \dot{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2} \quad (1)$$

$$MAPE(y, \dot{y}) = 100(\frac{1}{n}\sum_{i=1}^{n}\frac{|y_i - \dot{y}_i|}{y_i}) \quad (2)$$

$$RMSPE(y, \dot{y}) = 100\sqrt{\frac{1}{n}\sum_{i=1}^{n}(\frac{y_i - \dot{y}_i}{y_i})^2} \quad (3)$$

Here, y is defined as the vector of actual values, $\dot{y}_i$ is the vector of predicted values and $\bar{y}$ is the mean of actual values. Actual values refer to recorded measurements of indoor air temperature. Combining these error metrics yields promising insights into the model performance. In their literature review, Østergård et al. (2018) showed how R^2 values are often presented for a variety of different building SMs and allow an easy comparison between them despite the diversity of the problems. Therefore, it can be considered a metric of high importance as higher values indicate the closeness of fit and overall performance. The other errors are relative metrics where the RMSPE is more drastically affected by small, large errors than MAPE (Westermann et al. (2020)).

Zone Quantity Impact

Preliminary sensor data was conducted to eventually be used to predict the home's parameters. Unfortunately, the data gathering encountered an issue where some sensors' batteries were depleted faster than anticipated resulting in severe data gaps for many of the sensors. Time periods still remained where a few sensors were able to provide a sufficient amount of data. To anticipate the required reduction in the number of zones it was insightful to examine the impact of training the model for a smaller number of thermal zones. These zones included two on the first floor, one on the second and the attic. The binary heating column was also removed to investigate performance without it; therefore only 4 zones were used.

Results and Discussion

Training was originally completed with 100 epochs, which took roughly 3 minutes for each, while using a default Batch Size of 32. Early stopping was also included to prevent overfitting. Early results suggested that training for 1 parameter output typically yielded better results than training for all parameters at once (Table 2). Exceptions do remain (such as the R^2 for Zone Infiltration Flow Rate), but error metric scores more often tended to show a notable decrease when trained for 1 parameter instead of all.

Table 2: R^2 error metric results for training on all parameters at once and individually for the month of December.

	Train		Test	
Parameter	**Group**	**Individual**	**Group**	**Individual**
North Axis	0.795	-	0.755	-
Wall Insulation Conductivity	0.983	0.996	0.981	0.995
Roof Insulation Conductivity	0.976	0.984	0.969	0.979
Foundation Concrete Thickness	0.268	-	0.194	-
Window Conductivity	0.000	-	-0.004	-
Lighting Energy	0.426	-	0.462	-
Electric Equipment Energy	0.482	0.555	0.497	0.439
Personnel	0.560	-	0.532	-
Zone Ventilation Flow Rate	0.487	-	0.409	-
Zone Infiltration Flow Rate	0.972	0.958	0.968	0.959

Further investigation should be conducted by varying the number of epochs and batch size to examine the significance of the differences. Due to these perceived difference in errors, 4 parameters (highlighted in Table 2) were examined for the rest of the project while the others remained varied but not trained for, thereby creating noise to an extent. The choice of these 4 parameters was to capture those with the highest potential performance in terms of error metrics and one with poorer prediction accuracy.

Results on Synthetic Data

One disadvantage of switching from training on all outputs at once to only one is that it involves the creation of multiple ISMs (one for each time period and parameter). This slows down training significantly as each ISM requires its own training to be completed. To account for this disadvantage, the 4 parameters were trained at once instead of separately. The number of epochs was also increased from 100 to 150, with early stopping still enabled.

While the initial results in Table 2 do not seem quite promising, the overall analysis identifies the sensitivities of individual parameters on temperature. Henceforth, the ISM can easily identify highly influential parameters (such as Wall Insulation Conductivity) compared to those that only make minor adjustments in temperature. This is perhaps unsurprising, as thermal insulation and infiltration are expected to be a significant influence on energy demand (Zhiqiang John Zhai (2019)), hence it can be expected that variations among values can cause larger temperature fluctuations (i.e., more energy would be needed to provide a comfortable temperature which is most impacted by these parameters). This would thereby imply that the method may be best suited for determining envelope characteristics. Results for the ISMs for each of the 4 parameters and their respective time periods are shown in Table 3.

Table 3 indicates strong performance for ISMs to determine exact building parameter values. Both the insulation conductivities and infiltration flow rate showed high performance overall with the R^2 frequently above 95%. Even when trained individually, Electric Equipment Energy (a previous poor performing parameter), still displayed disappointing performance. It should not be understated though that the ISMs are still able to provide a somewhat accurate, though not ideal, estimate for it. This would suggest that an ISM, depending on the user's requirements, may still be adequate in providing at least an idea of where the building parameter value lies.

An interesting observation can be made that the R^2 values in Table 3 tend to be slightly lower than those in Table 2. While this difference is minimal, it remains surprising. A possible explanation is that while the train and test sets were taken from the same pool of data, the distribution of the data into the two sets was different (i.e. some data used for training with 15 parameters were used for testing in the case of 4 parameters). This may suggest that the training set used when training for 15 parameters was better suited than the set with 4 parameters. Larger data sets will be explored in the future to determine if improvements can occur. Furthermore, the improvement of reducing the number of parameters may not be seen, as the discrepancy between the R^2 values for the Electric Equipment Energy parameter suggests that overfitting might be occurring. As early stopping is being applied, a possibility exists that training is being ended too early in an attempt to limit the overfitting of this parameter, while the other parameters still require longer training periods.

Lastly, it is worth mentioning the performance of the North Axis parameter shown in Table 2. One concern of applying the error metrics to this parameter is that the values are in degrees of a circle (e.g., 0-360°). Therefore, as an example, while values of 5° and 355° may not actually be significantly far apart, the error metric calculates the difference as being 350° thereby producing significant error. Future work will explore better methods of performance evaluation for this parameter.

Comparison between 4 and 15 zones

The comparison of results between ISMs trained on 15 zones and those trained on 4 zones are shown in Table 4.

Unsurprisingly, the ISMs perform worse when only trained on 4 sets of temperature data instead of 15. Differences are mostly minimal for the R^2 values for the insulation conductivities, however a large rise in both the RMSPE and MAPE values are notable. Additionally, the Zone Infiltration Flow Rate performs significantly worse overall with the R^2 going below 90%. Interestingly, the Electric Equipment Energy parameter achieves a slightly higher performance with reduced zones, however the R^2 is still much poorer in comparison to the other parameters. The reduction in zones would appear to remove key informative locations, thereby reducing the ability of the model to train effectively. This is especially notable for the Roof Insulation Conductivity parameter as zones on the main floor of the building (not directly touching the roof) are much less influenced by the conductivity of the roof insulation. One of the remaining 4 zones is the attic, hence the ISM performance is still solid as it is heavily influenced by the parameter, however this still demonstrates that caution needs to be applied when selecting a subset of zones to train on.

Future Work

Future work largely pertains to eventually being able to perform predictions with real data. Significant improvements can still be made to the current building simulation model. While EnergyPlus has shortcomings when replicating real building energy simulation performance, an inadequate representation is often much more a result of poor input data (Edwards et al. (2017)). For this reason, it is imperative that the building simulation model is designed to mimic the current home as accurately as possible. In our case, inaccuracies can be largely attributed both to the model inputs that were not varied (such as setpoints) and additional input (e.g., the weather file). Temperature setpoints should be adjusted such that

Table 3: Error metric results for the synthetic data train and test sets. Colouring is applied separately for each error metric. An upper bound of 30% was only used for colouring and both RMSPE and MAPE can go higher.

Training	Wall Insulation Conductivity			Roof Insulation Conductivity			Electric Equipment Energy			Zone Infiltration Flow Rate		
	R^2	RMSPE	MAPE	R^2	RMSPE	MAPE	R^2	RMSPE	MAPE	R^2	RMSPE	MAPE
March	0.976	10.263	7.074	0.968	11.729	7.965	0.598	10.183	8.18	0.982	6.264	4.573
June	0.95	13.029	9.572	0.969	11.844	8.127	0.444	12.832	10.027	0.976	9.413	6.171
Sept	0.96	15.182	10.566	0.967	16.879	10.163	0.563	10.528	8.518	0.978	10.152	6.598
Dec	0.973	11.926	7.904	0.942	14.142	10.293	0.553	11.324	8.878	0.977	9.224	6.244
Testing	**Wall Insulation Conductivity**			**Roof Insulation Conductivity**			**Electric Equipment Energy**			**Zone Infiltration Flow Rate**		
	R^2	RMSPE	MAPE	R^2	RMSPE	MAPE	R^2	RMSPE	MAPE	R^2	RMSPE	MAPE
March	0.974	10.796	7.383	0.968	13.213	8.888	0.477	11.187	9.01	0.977	6.679	4.999
June	0.949	13.375	9.593	0.97	12.173	8.205	0.361	13.327	10.399	0.974	9.732	6.259
Sept	0.958	13.842	10.008	0.962	19.352	11.501	0.462	11.441	9.097	0.976	9.975	6.713
Dec	0.97	11.781	7.97	0.945	13.379	10.134	0.447	12.185	9.494	0.976	8.931	6.205
Legend	R^2	0										1
	RMSPE & MAPE	0										30

Table 4: Comparison between results of the previous 15 zone ISMs and 4 zone ISMs.

Parameters	15 Zone Test Set	4 Zone Test Set
Wall Insulation Conductivity	R^2: 0.970, RMSPE: 11.781, MAPE: 7.97	R^2: 0.940, RMSPE: 19.901, MAPE: 12.094
Roof Insulation Conductivity	R^2: 0.945, RMSPE: 13.379, MAPE: 10.134	R^2: 0.935, RMSPE: 27.172, MAPE: 15.167
Electric Equipment Energy	R^2: 0.447, RMSPE: 12.185, MAPE: 9.494	R^2: 0.458, RMSPE: 11.588, MAPE: 9.310
Zone Infiltration Flow Rate	R^2: 0.976, RMSPE: 8.931, MAPE: 6.205	R^2: 0.892, RMSPE: 15.184, MAPE: 11.364

the minimum and maximum zone temperatures are an accurate reflection of those currently in the home. Furthermore, the weather file currently used for simulations does not contain the exact weather features of the time period when real data was recorded. Lastly, as mentioned previously, larger train and test sets will be explored as well as better methods of evaluation for the North Axis parameter

While future work is promising, it still unfortunately remains difficult for the ISM to estimate the exact values of the home's inner components. Additionally, as the exact parameter values are unknown, a comparison to the actual cannot be made to what was predicted. However being able to predict in expected bounds of the parameters (i.e. those in Table 1) could still suggest the method is an effective method of parameter prediction.

Conclusions

This paper presents the development and use of a Inverse Surrogate Model designed to accept building temperature time series data as input and accurately predict building parameter values. ISMs could serve as a cost effective alternative method of building parameter identification. This would assist home owners with relevant information to help determine if a retrofit is justifiable. When trained and tested on synthetic data, the model showed high performance with the insulation conductivity parameters and infiltration flow rate. Due to limitations when trying to collect real data with physical sensors, the model was retrained with only 4 zones worth of data. This produced similar results, however there was a notable decrease in performance among the conductivity and infiltration flow rate parameters.

Predicting with real sensor data was not completed at this stage of the project. As the results of some of the ISMs on synthetic data are promising however, future work is planned with a focus of achieving strong performance on real data. This will include creating a more appropriate weather file and performing slight model adjustments and editing temperature setpoints for each individual zone/room to ensure that the temperatures reach a more correct default.

References

Bamdad, K., M. E. Cholette, S. Omrani, and J. Bell (2021). Future energy-optimised buildings — addressing the impact of climate change on buildings. *Energy and Buildings 232*.

Edwards, R. E., J. New, B. Cui, and J. Dong (2017). Constructing large scale surrogate models from big data and artificial intelligence. *Applied energy 202*, 685–699.

Lawrence, C. R., R. Richman, M. Kordjamshidi, and C. Skarupa (2021). Application of surrogate modelling to improve the thermal performance of single-family homes through archetype development. *Energy and Buildings 237*.

P.Angelov, P. and X. Gu (2019). *Empirical Approach to Machine Learning*. Springer.

Pedregosa, F., G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel,

P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research 12*, 2825–2830.

Shields, M. D. and J. Zhang (2015). The generalization of latin hypercube sampling. *Reliability Engineering & System Safety 148*, 96–108.

Statistics Canada (2016). Census profile, 2016 census - victoria [census metropolitan area], british columbia and british columbia [province]. `https://www12.statcan.gc.ca/census-recensement/2016/dp-pd/prof/details/page.cfm?Lang=E&Geo1=CMACA&Code1=935&Geo2=PR&Code2=59&Data=Count&SearchText=victoria&SearchType=Begins&SearchPR=01&B1=All&TABID=1`. Accessed: 2022-02-10.

Urge-Vorsatz, D., K. Petrichenko, M. Staniec, and J. Eom (2013). Energy use in buildings in a long-term perspective. *Current Opinion in Environmental Sustainability 5*(2), 141–151.

United States Department of Energy (2022, March). *EnergyPlus Version 22.1.0 Documentation: Input Output Reference*.

Westermann, P. and R. Evins (2019). Surrogate modelling for sustainable building design – a review. *Energy and Buildings 198*, 170–186.

Westermann, P., M. Welzel, and R. Evins (2020). Using a deep temporal convolutional network as a building energy surrogate model that spans multiple climate zones. *Applied Energy 278*.

Zhiqiang John Zhai, J. M. H. (2019). Implications of climate changes to building energy and design. *Sustainable Cities and Sociey 44*, 511–519.

Østergård, T., R. L. Jensen, and S. E. Maagaard (2018). A comparison of six metamodeling techniques applied to building performance simulations. *Applied Energy 211*, 89–103.

6.3 Appendix C Inverse Surrogate Modelling to Determine Thermal Characteristics of Buildings

Inverse Surrogate Modelling to Determine Thermal Characteristics of Buildings

Liam Jowett-Lockwood[1,2], Ralph Evins[1,2]
[1]Energy in Cities group, Department of Civil Engineering, University of Victoria, BC, Canada
[2]Institute for Integrated Energy Systems, University of Victoria, BC, Canada

Abstract

Planning building energy retrofits effectively requires knowledge of the current state of the building envelope, which is often lacking in practice. This study examines the usage of an Inverse Surrogate Model (ISM) in the form of a Convolutional Neural Network (CNN) for the purposes of determining various building parameters, such as the wall insulation conductivity. The training dataset was based on EnergyPlus simulations from an existing building and consisted of temperature time series data and their corresponding varied parameters. The ISM performance is assessed with a variety of common statistical error metrics and results indicated strong performance with most parameters.

Highlights

- ISM uses a CNN to determine building characteristics to assist in retrofit planning.
- Tested for varying time intervals, including monthly, weekly and daily.
- Results showed high prediction performance.

Introduction

The majority of homes in Canada are decades old and turnover in the building stock is low. From the 2021 Census of Population (based on 25% sample data), in British Columbia 39% of homes were built prior to 1981 and 12% of homes were built prior to 1961 (Statistics Canada (2021)). A significant energy reduction can be obtained by applying proper retrofit strategies alongside innovative construction techniques and new technologies (Urge-Vorsatz et al. (2013)). However, before attempting a retrofit, it is important to carefully assess the performance of the building. Doing retrofits haphazardly by not fully understanding the existing structure can lead to disappointing results. This can result in situations where the completed retrofit is insufficient in meeting its intent and that the necessary retrofit measures to achieve a desired energy performance goal is too expensive (Urge-Vorsatz et al. (2013)). Additionally, more consideration should be applied as, while smaller than a complete new construction, there is still an environmental consequence of increasing the embodied energy of conducting a retrofit (Shadram et al. (2020)). However, older buildings may be missing important key information, such as the U-value of their exterior wall insulation or the infiltration flow rate. It is important to develop non-intrusive and easily applied methods of determining these important characteristics. This study investigates the use of a machine learning method in concert with building energy simulations to provide a method of estimating unknown building parameters from temperature time-series. Such data could be obtained from smart thermostats or inexpensive sensor deployments.

Inverse Modelling

Computational methods of building parameter identification offer notable advantages over their non-computational counterparts. Notably, non-computational methods can include tests (such as a blower door test to determine the infiltration flow rate) which can be expensive and require intrusive access to the building (Edwards et al. (2017)). As an alternative, inverse modelling presents itself as a computational approach to determine building parameters. There are various types of inverse models that exist. One of the simplest approaches for inverse modelling is via calibrated simulation models. These models make use of the classical forward approach that involves the development of a building energy model with physics-based simulation software such as EnergyPlus (USDOE (2023)) or IES-VE (IES (2023)). These building energy models incorporate physical input parameters that will produce, among others, energy usage outputs when run. The inverse model itself will tune the building parameters in the building simulation model until the outputs match those desired with a relative degree of uncertainty. This approach has two major hindrances. Firstly, the detailed building information needed may be incomplete or difficult to obtain and then secondly, the development and calibration process is not only user intensive, but also computationally slow (Zhang et al. (2015)). An alternative to calibrated simulation models are data-driven models which (with enough data)

can form statistical relationships between given related inputs and outputs. The most common form of these models are regression type models, which rely on the assumption that nonlinear energy use or additional output data can be adequately captured by building parameters with regression (Zhang et al. (2015)). Recently the usage of machine learning, in particular Artificial Neural Networks (ANNs), for inverse modelling has gained significant traction. This has been aided in part by the development of a Surrogate Model (SM) as an alternative to traditional building energy simulation.

Surrogate Modelling

A SM is a machine learning model that, once trained, can effectively predict the outputs that would be produced from an intensive computational model (Westermann and Evins (2019)). One of the strongest benefits of a SM is how the predictions are almost instantaneous when compared to traditional building energy simulation modelling. In the context of buildings, a building SM is traditionally trained on inputs to a building simulation model (such as EnergyPlus) to predict the outputs from the building simulation model. Inputs in this case traditionally consist of building parameters, such as the exterior wall insulation U-value and infiltration flow rate, and outputs like aggregate energy usage. It has been shown that the usage of SM is widespread with successful applications including early conceptual design, sensitivity analysis, uncertainty analysis and optimization (Westermann and Evins (2019)).

One of the limitations of developing a SM however, is the acquisition of training data. Obtaining enough synthetic simulation data necessary to train the SM potentially requires a notably large amount of simulation runs. While providing a time estimate on generating the training data is largely dependent on the computing hardware used, due to the nature of a SM, predictions produced remain almost instantaneous even without extensive high performing computer hardware. While the acquisition of training data may seem like a substantial hindrance as the extensive number of simulations would appear to deem the development of a SM nonsensical, if a SM can generalize well (i.e., be effectively applied to other projects and structures) then this shortcoming is accounted for (Westermann and Evins (2019)). This has the possibility to greatly extend the applicability of a SM and therefore, further increase its advantage over traditional building simulation, as the process of reacquiring training data is now removed.

Inverse Surrogate Modelling

While SMs are an effective means of predicting the outputs of a building simulation model, an Inverse Surrogate Model (ISM) suggests doing the inverse. Whereas a SM is trained on building energy simulation inputs to predict energy simulation outputs, an ISM would be trained on said outputs to instead predict the inputs. While a complete set of training data would still be required, doing so allows another alternative to building parameter identification and inverse modelling. If successful, an ISM has the potential to produce accurate predictions of various building characteristics with only simulation output data, such as an energy or temperature time-series. If applied to real data, this offers a potential solution to being able to make accurate, inexpensive and non-intrusive assessments of the characteristics of one's home.

Literature Review

The usage of inverse modelling has largely been restricted to that of determining an energy use baseline and potential savings from the application of retrofits (Burak Gunay et al. (2021)). More seldom done however, is the application of determining physical building parameter inputs via inverse modelling. Ko, et. al, used an inverse modelling approach with transfer learning to enable an inverse model to accurately predict the nominal cooling coefficient of performance, wall U-value and lighting power density (Ko and Park (2021)). This transfer learning approach first trained an inverse model with data produced from 10,000 EnergyPlus simulations, which was then trained on non-synthetic data from 49 buildings to then estimate the three parameters of 12 remaining buildings in the dataset. Burak et al. developed a method to characterize envelope thermal transmittance and air infiltration using non-synthetic space heating load and building automation data (Burak Gunay et al. (2021)). This multi-step process first examined the parameters for the weekly ventilation schedule, then examines the air permanence by modelling the rate decrease of the average CO2 concentration and then lastly, either steady-state, single-node transient or multi-node transient change point models are trained to determine the thermal transmittance. The results indicated that all three models formed in the last step produced relatively similar performance with transient models providing a slight edge. Romallo-González et al. investigated inverse modelling with Lumped Parameter Models for 6 existing buildings and 1000 simulated building models (Ramallo-González et al. (2018)). Their work examined the ability to estimate the heat transfer coefficient, along with internal temperatures, electricity use, and CO_2 concentration. Their findings found that reasonable estimates can be achieved during the wintertime period, suggesting that reasonable thermal properties estimates can be obtained with a small number of sensors. Most closely aligned with this study is work recently completed by Herbinger et al. (2023). Their work involved the use of an ANN for the purposes of determining building parameters. The authors demonstrated high performance of their model with predicting 14 building parameters with

hourly energy data as an input. Due to the nature of their ANN however, they used a staggering 16384 neurons in order to handle their energy time-series data as input.

Areas of Improvement

The purpose of this study is to expand on the usage of inverse modelling for building envelope parameter identification with an ISM. While most ANNs currently used in literature are simplistic, we instead focus on successfully implementing a deep learning Convolutional Neural Network (CNN) because of their demonstrated high performance with time series inputs (Westermann et al. (2020)). CNNs utilize convolutional kernels among others to extract key features from a large quantity of inputs. For this study, the feature extraction ability of CNNs are appropriate to handle the multiple extensive temperature time series inputs, whereas a traditional ANN would potentially require a significant amount of memory usage and training time (Goodfellow et al. (2016)). Recurrent Neural Networks (RNNs) are also worth considering in this regard. RNNs are designed specifically to handle a series of input sequentially and can produce outputs either once the sequence has been completed or at every timestep. Historically the recurrent nature of RNNs have made them highly susceptible to the vanishing gradient problem when tasked with long-term dependencies, thereby reducing effectiveness in these scenarios (Goodfellow et al. (2016)). Later versions of RNNs, namely Long Short-Term Memory (LSTM) networks, attempt to rectify this issue, whereas the specific residual nature of the CNN used in this study also alleviates this problem. To the best of the authors understanding, we are the first to use a CNN as an ISM for the purposes of building envelope parameter prediction.

Eventually, this research aims to investigate the performance of the ISM when making predictions with non-synthetic collected sensor data, however that is currently out of scope for this study. Additionally, this study examines the extent of how much data is needed to produce accurate observations. This includes not only the duration of data required (i.e., monthly, weekly, and daily), but also the number of different recordings. This is beneficial for the collection of real data, as demonstrating that only small quantities of data are needed, would enable accurate predictions off of a short duration of sensor deployment.

It should be noted that an earlier version of this research by the authors was presented at the eSim 2022 conference hosted by Carleton University in Ottawa Canada (Carleton (2023)). That research followed a similar methodology to this study with the main difference being that it explored the effect of seasonality with the ISM when trained on a different month's worth of data (Jowett-Lockwood and Evins (2022)). The study now presented in this paper places emphasis on the significance of the ISM in making accurate predictions with different time resolutions. Other changes for this study include the creation and usage of a time-relevant weather file, a larger number of parameters predicted, changes in error metrics and others.

Methodology

This project aims to explore the effectiveness of an ISM in its ability to produce accurate predictions of building parameters based on synthetic data for different time resolutions. The study is largely composed of two parts. Firstly, the building simulation model derivation and data configuration. Secondly, the CNN development and data evaluation. The overall framework of the study is illustrated in Figure 1.

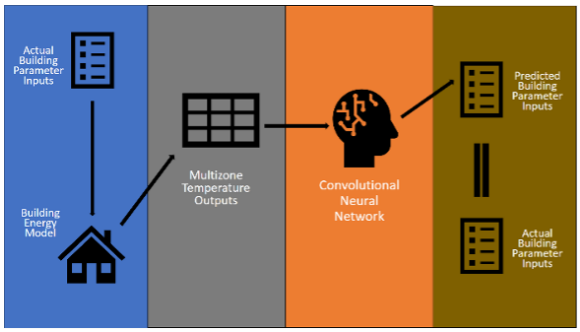


Figure 1: Study methodology.

Synthetic Data Generation

The data generation was an imperative part of the study, as it involved the creation of data for training. Even though information about the existing structure is lacking, it is still necessary to design the model such that it still produces a reasonably accurate reflection of the building's performance. Doing so will assist in aiding the ISM in eventually predicting with actual data. To create the building simulation model, geometry characteristics were used alongside existing information about the home as well as the development and inclusion of an accurate weather file for the time period of data recording and necessary formatting and pre-processing.

Home Characteristics

The building used for this study was a residential single-family home built in the 1930s in Victoria, British Columbia, Canada from which real data was taken from for use in an additional study. The geometry of the model is shown in Figure 2 with simulations being computed with EnergyPlus. For the simulation model, 14 zones were chosen to roughly match rooms of the home. These included areas such as the attic and hallways. In total 7 zones are present on the first floor, 5 for the second floor and 2 additional zones represent the basement and attic. Windows were treated as uniform in terms of characteristics to make the model parameterization simpler. It was noted that wall insulation was less conductive

on the second floor, therefore the parameterization of the wall insulation was split.
No comprehensive HVAC system was established for the home, rather mechanical heating and cooling were provided via the Ideal Air Loads System Object. To constrain the system from too quickly meeting temperature setpoint changes (as in reality), the Maximum Heating Air Flow Rate parameter was parameterized. Due to the uncertainty between zones, the infiltration and ventilation were also defined simply, such that their flow rates were simply parameterized with no additional formulas.

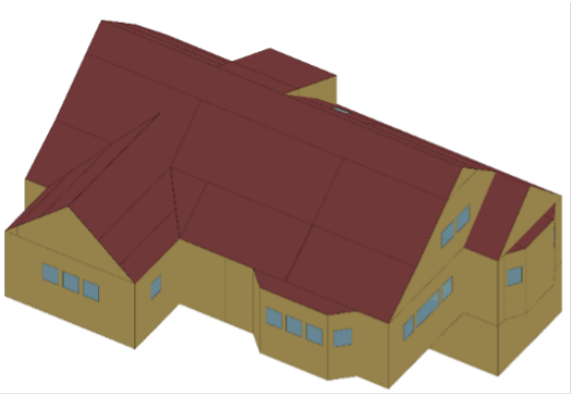


Figure 2: SketchUP model based on an existing home.

Weather File

To ensure that the simulation outputs were accurate, it was imperative for the simulations to include actual weather from the time period of data collection. The process for developing the necessary weather file largely followed the approach taken by Evins et al. (2020). The collection of data retrieved was taken from the School-Based Weather Station Network on Vancouver Island. The network is a collection of weather stations on the island for educational purposes and collects atmospheric temperature and humidity, UV Index, solar radiation, wind speed and direction and atmospheric pressure (Evins et al. (2020)). Fortunately, because of the numerous schools within the network, weather data was available at a location only about a kilometer away from the home. Cloud cover was unfortunately not recorded at that location. As this was required for the weather file, qualitative non-numerical cloud cover observations (e.g., cloudy, mostly cloudy, clear, etc.) were taken from the data collected by the Victoria International Airport (Government of Canada (2023)). As the weather file required cloud cover values in a numerical range (e.g., 0 to 8 with 8 being completely overcast and 0 being clear skies) assumptions were made about the perceived value based on how cloudy the reported observation was.

Data Organization

To generate multiple simulation runs, different parameters were modified, see Table 1. It is expected that only parameters that have significant influence

Table 1: Varied parameters used for the development of training data.

Parameter	Unit	Range
Upper Wall Insulation Conductivity	W/mK	0.01-0.1
Lower Wall Insulation Conductivity	W/mK	0.01-0.1
Roof Insulation Conductivity	W/mK	0.01-0.1
Floor Concrete Thickness	m	0.05-0.2
Glass Conductivity	W/mK	0.005-0.02
Attic Conductivity	W/mK	0.01-0.1
Lighting Energy	W/m^2	5-10
Electric Equipment Energy	W/m^2	5-10
People per area	$People/m^2$	0.015-0.05
Ventilation Flow Rate	L/sPerson	5-10
Infiltration Flow Rate	L/m^2s	0.1-1
Maximum Heating Air Flow Rate	L/s	1-10

on the temperature are noteworthy to train for. Those which do not have a major influence on the temperature are likely to be difficult to predict and most changes will go unobserved in the data. Therefore, only envelope characteristics were parameters which were trained for. All parameters remained varied however, with the non-trained parameters functioning as additional noise. The parameters themselves were chosen based on previous research (Westermann et al. (2020)). Parameters highlighted in green were those which the ISM was trained to predict while others in red were included to see the ISM's ability to handle additional noise. The determination of which parameters to make predictions for and those not to is based on previous research (Jowett-Lockwood and Evins (2022)). The ranges of potential values were chosen based on roughly where it is anticipated that the value lies for the actual structure, while the extents were decided by engineering judgement. The sampling itself was completed with Latin Hypercube Sampling (LHS), which is frequently used and highly regarded for its ability to effectively reduce the sampling variance for a large variety of applications (Shields and Zhang (2016)).
For data production, a total of 5000 simulations were produced to be used as training data. After each simulation run, the data was organized into a CSV file via a Python script. Temperature values were recorded for each zone in 10-minute intervals resulting in a total of 52560 (10-minute intervals in a year) values, but in practice the actual amount of values were much smaller as only monthly, weekly and daily data time periods were examined. Alongside the 14 zones, an additional binary column for heat demand was included. This was treated as a 15th zone in

the data with values of 1 indicating heat was being applied to the home and a value of 0 indicating otherwise. The intention of this column was to help provide the ISM with additional information for when temperature fluctuations alone may be insufficient.
Additionally, prior to data being fed to the ISM, it was important that the data was standardized to resemble standard normally distributed data. Standardization works by re-scaling the values of features so that they fit a desirable range. Doing so, ideally lowers the differences between the feature values so that their values are more comparable to each other (Angelov and Gu (2019)).

Convolutional Neural Network Structure

CNNs can simply be composed of convolutional layers that are often used in conjunction with nonlinear activation and pooling layers (Goodfellow et al. (2016)). While an activation layer simply accepts the input and applies the activation function (in this case ReLU), a pooling layer is more sophisticated. Pooling layers differ from dense layers in that they compute either the average or max value of a subset of nodes in order to effectively lower the size of the data to be outputted to the next layer. It is possible for CNNs to often follow a structure of several iterations of these layers until eventually finishing.
The structure of the CNN used for the ISM in this study is shown in Figure 3. This residual CNN structure is based heavily on the CNN deployed by Westermann et al. (2020). The model is predominately composed of three blocks of layers. Within each block is a one-dimensional convolutional, batch normalization and Rectified Linear Unit (ReLU) activation layer with the later two replacing the aforementioned pooling layers. A global pooling layer remains near the end of the model alongside a L2 regularization layer. Most importantly, the residual network contains residual shortcuts connecting the ends of each block together. Their purpose is to prevent the vanishing gradient from occurring and thereby improving training ability (Westermann et al. (2020)). The reader is referred to their study for a more comprehensive breakdown of the model characteristics. Aside from layer resizing, the model's most significant difference is that there is no concatenation of additional features at the end of the model, as it was not necessary in this study.

Error Metrics

Analyzing error metric values was the primary method of model assessment. Error metric equations take in both arrays of true and predicted values to determine the model's predictive effectiveness. The selection of error metrics needs to be carefully chosen as different error metrics can indicate different characteristics about the model's performance. For this study we examined the performance with the Coefficient of Determination (R^2), Mean Absolute Percentage Error (MAPE) and Root Mean Square Error (RMSE) defined below:

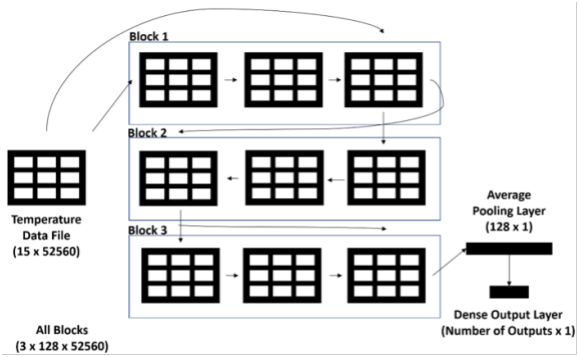


Figure 3: Convolutional Neural Network Structure.

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

$$MAPE(y,\hat{y}) = 100(\frac{1}{n}\sum_{i=1}^{n}\frac{|y_i - \hat{y}_i|}{y_i}) \quad (2)$$

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

The values of y refer to the actual data (with $\overline{y}$ being the mean), while $\hat{y}$ refers to the predicted values from the model.

The Coefficient of Determination is a frequently used unit less metric in machine learning with values typically ranging from 0 to 1 with values closer to 1 indicating higher performance. A strong benefit of having an upper bound for this metric is the ease in which it can be used to compare model accuracies to diverse problems across multiple studies (Østergård et al. (2018)). This is advantageous for this study, as it helps put into perspective the prediction performance of the ISM.
One of the most significant advantages of Mean Absolute Error (MAE) and RMSE is their ease of understanding. The most significant difference between the two-error metrics is the quadratic term in the RMSE equation. This increases the significance of small, but large errors (outliers) on the result (Westermann et al. (2020)). With the inclusion of RMSE, it was decided to forgo using Mean Absolute Error (MAE) in favor of MAPE. MAPE retains the simplicity and ease of understanding of MAE, while offering the difference of being a relative metric and scale independent.
In the event of significantly large outliers, there exists a possibility that the error metric may report incredibly poor performance even though the offending samples may be a very small amount. It has been decided that regardless of the unit and parameter, predictions less than 0 and greater than 1 will be omitted from the error metric calculation. Similar approaches have

Table 2: Monthly error metric results.

MONTHLY	Glazing Conductivity	Infiltration Flow Rate	Attic Insulation Conductivity	Upper Wall Insulation Conductivity	Roof Insulation Conductivity	Lower Wall Insulation Conductivity	Scale
R^2 - Train 15	0.967	0.997	0.962	0.976	0.994	0.986	0.000
R^2 - Test 15	0.950	0.997	0.955	0.966	0.993	0.981	0.333
R^2 - Train 4	0.826	0.994	0.010	0.935	0.994	0.954	0.666
R^2 - Test 4	0.540	0.993	-0.013	0.841	0.992	0.938	1.000
MAPE - Train 15	5.716	1.962	13.357	8.201	2.815	4.761	0.000
MAPE - Test 15	6.745	2.143	13.222	8.821	3.270	5.014	23.000
MAPE - Train 4	11.368	3.617	67.762	12.463	4.422	12.529	46.000
MAPE - Test 4	18.747	4.240	67.828	20.422	4.789	12.494	70.000
RMSE - Train 15	0.007890	0.000014	0.005050	0.004015	0.001970	0.003103	0.000000
RMSE - Test 15	0.000970	0.000015	0.005501	0.004780	0.002188	0.003507	0.033333
RMSE - Train 4	0.001806	0.000020	0.025889	0.006641	0.001995	0.005589	0.066666
RMSE - Test 4	0.002941	0.000022	0.025977	0.010284	0.002333	0.006392	0.100000

Table 3: Weekly error metric results.

WEEKLY	Glazing Conductivity	Infiltration Flow Rate	Attic Insulation Conductivity	Upper Wall Insulation Conductivity	Roof Insulation Conductivity	Lower Wall Insulation Conductivity	Scale
R^2 - Train 15	0.915	0.999	0.975	0.970	0.997	0.995	0.000
R^2 - Test 15	0.884	0.999	0.951	0.949	0.996	0.992	0.333
R^2 - Train 4	0.924	0.994	0.000	0.907	0.990	0.989	0.666
R^2 - Test 4	0.810	0.992	-0.017	0.816	0.987	0.979	1.000
MAPE - Train 15	7.265	1.384	7.134	9.286	2.470	3.909	0.000
MAPE - Test 15	8.788	1.615	9.494	11.363	2.678	4.224	23.000
MAPE - Train 4	8.013	3.344	68.079	14.593	5.419	4.951	46.000
MAPE - Test 4	12.623	4.245	68.059	20.512	6.120	6.591	70.000
RMSE - Train 15	0.001129	0.000008	0.003882	0.004429	0.001305	0.001879	0.000000
RMSE - Test 15	0.001361	0.000009	0.005465	0.005792	0.001464	0.002240	0.033333
RMSE - Train 4	0.001190	0.000019	0.026022	0.007873	0.002482	0.002670	0.066666
RMSE - Test 4	0.001861	0.000023	0.026026	0.010887	0.002819	0.003734	0.100000

been taken in literature with predictions where the actual value is negative and are then omitted (Kim and Kim (2016)).

Results

Testing and evaluation were broken down between the month-long period, which included evaluation for the entire month itself, the 4 weeks and each day composed in it. Evaluations with 4 zones were conducted for daily, weekly and the entire month time periods, however evaluations with the full 15 zones were not conducted daily. This was deemed unnecessary as results with 4 zones proved to show high performance. Results provided in Table 3 and Table 4 represent the average error metric values across all weeks and days respectively. Additionally, while occasional significantly large outliers were detected and removed, they made up a minuscule portion of the predictions (often less than 1%).

Monthly Synthetic Data Results

Table 2 shows R^2, MAPE and RMSE monthly performance on purely synthetic data with 15 zones. Performance remains high across all parameters with most observations having R^2 scores above 90%. This remains the case after the transition to 4 zones, however both the glazing and the attic insulation conductivity have notable performance drops. Overfitting could be a potential reason for the discrepancy between the train and test performance for the glazing, however, both the train and test performance for the attic is exceptionally poor. The cause for which is most likely attributed to the lack of representation these parameters have on the 4 recorded zones. Especially in the case of the attic, its influence on zones that are

Table 4: Daily error metric results.

DAILY	Glazing Conductivity	Infiltration Flow Rate	Attic Insulation Conductivity	Upper Wall Insulation Conductivity	Roof Insulation Conductivity	Lower Wall Insulation Conductivity	Scale
R^2 - Train 4	0.875	0.992	0.064	0.866	0.988	0.981	0.000
R^2 - Test 4	0.777	0.988	-0.003	0.738	0.979	0.965	1.000
MAPE - Train 4	9.373	4.062	64.058	17.186	4.205	6.695	0.000
MAPE - Test 4	13.077	5.225	65.746	24.109	5.244	8.270	70.000
RMSE - Train 4	0.083442	0.000022	0.025068	0.008980	0.002283	0.003498	0.000000
RMSE - Test 4	0.001911	0.000028	0.025837	0.012606	0.002915	0.004691	0.100000

sufficiently far away, remains minimal. Furthermore, the decrease in available glazing across 4 zones would seem to further imply that lack of influence being the key reason for the performance decrease.

Weekly Synthetic Data Results

Weekly results are provided in Table 3. What is notably apparent is how results are similar to those in Table 2, thereby suggesting that moving to weekly time periods is largely insignificant in terms of performance. A surprising feature of the results in Table 3 is the increase in prediction performance with the glazing. Whereas with monthly results where there was a significant discrepancy with the train and test set performance with 4 zones, this difference is much smaller. It is difficult to determine the reasoning behind this change, however there is a possibility that the training set in this instance more closely resembles the testing set by coincidence, thereby enabling the ISM to more easily make accurate predictions.

Daily Synthetic Data Results

When compared to monthly and weekly results, daily results (Table 4) show high performance for most parameters, however, decreases in performance compared to longer time periods are noticeable. Notably, the upper wall insulation parameter performs particularly worse than before. Similar to glazing, an explanation for the lower performance of the model with upper wall insulation could be that, given the roof of the building, less surface area and zones contain the parameter. This implies that the switch to 4 zones remains largely impactful for certain parameters and the decrease in the amount of time provided is an eventual hindrance.

When examining the trends shown in Figure 4, it can be observed that performance decreases tend to be consistent among most of the parameters. Notably, with a few exceptions, the Glazing Conductivity and Upper Wall Insulation Conductivity, tend to perform poorly and well on the same days. This would suggest that not only is duration an important aspect of training in this case, but also the quality of the daily data provided. Certain days may exhibit more significant temperature fluctuations because of

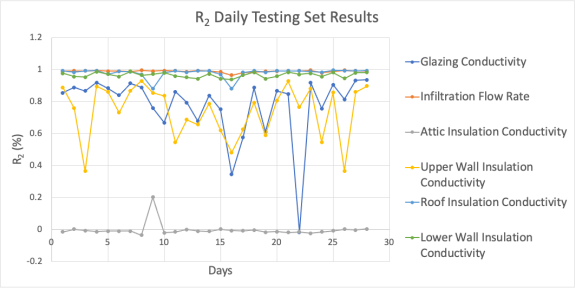


Figure 4: Daily R^2 values.

outside weather. As the zone temperature differences are more affected on these days than calmer days, the ISM is more likely able to understand the influence of the parameters on heat retention and thereby have increased prediction performance.

Future Work

Aside from eventually examining the ISM's performance on real data, the most significant area of improvement for handling synthetic data are more accurately performing predictions with only daily data. Examining the performance of other relevant ANN models, namely LSTM and other RNN variants, could provide insight into how effective the CNN is with lower quantities of data. Transformers are another ANN that have recently garnered much attention for its strong ability to handle language processing (Vaswani et al. (2017)), which would be another worthwhile comparison if it could be applied to the problem successfully. Additionally, examining the option of making the model "physics-informed" may alleviate training difficulty and improve performance for the daily time period. This could be completed by adjusting the loss function of the CNN, such that it is more guided in the appropriate direction.

Conclusions

This study examines the development of an ISM for the purposes of accurately determining building parameters based on only temperature measurements. This study examines the effectiveness of the ISM on synthetic data. The ISM itself is based on a CNN due to known success with time series inputs. 5,000 sim-

ulations were run to provide the training data, which was performed over monthly, weekly and daily time periods. Ultimately, the performance of the model on synthetic data was high for most parameters and remained quite high when tested for smaller time periods. Some parameters, namely the attic insulation and glazing conductivity, suffered performance decreases when switching from 15 zones to 4, which is anticipated to be due to the lack of representation these parameters had with only 4 zones.

References

Angelov, P. P. and X. Gu (2019). *Empirical Approach to Machine Learning* (1st ed. 2019. ed.). Studies in Computational Intelligence, 800. Springer International Publishing.

Burak Gunay, H., D. Darwazeh, S. Shillinglaw, and I. Wilton (2021). Remote characterization of envelope performance through inverse modelling with building automation system data. *Energy and buildings 240*, 110893–.

Carleton (2023). esim 2022. `https://carleton.ca/esim22/en_homepage/`. [Online; accessed 26-March-2023].

Edwards, R. E., J. New, L. E. Parker, B. Cui, and J. Dong (2017). Constructing large scale surrogate models from big data and artificial intelligence. *Applied energy 202*(C), 685–699.

Evins, R., R. Alexandra, E. Wiebe, M. Wood, and M. Eames (2020). The impact of local variations in a temperate maritime climate on building energy use. *Journal of building performance simulation 13*(2), 167–181.

Goodfellow, I., B. Yoshua, and C. Aaron (2016 - 2016). *Deep learning*. Adaptive computation and machine learning. The MIT Press.

Government of Canada (2023). Historical climate data. `https://climate.weather.gc.ca/`. [Online; accessed 25-February-2023].

Herbinger, F., C. Vandenhof, and M. Kummert (2023). Building energy model calibration using a surrogate neural network. *Energy and buildings 289*, 113057–.

IES (2023). Ve virtual environment. `https://www.iesve.com/software/virtual-environment`. [Online; accessed 25-February-2023].

Jowett-Lockwood, L. and R. Evins (2022). Using a convolutional neural network to determine the thermal characteristics of a building. In *Proceedings from eSim2022: Building Simulation Conference*. Ottawa (Canada), 21-24 June 2022.

Kim, S. and H. Kim (2016). A new metric of absolute percentage error for intermittent demand forecasts. *International journal of forecasting 32*(3), 669–679.

Ko, Y.-D. and C.-S. Park (2021). Parameter estimation of unknown properties using transfer learning from virtual to existing buildings. *14*(5), 503–514.

Ramallo-González, A. P., M. Brown, E. Gabe-Thomas, T. Lovett, and D. A. Coley (2018). The reliability of inverse modelling for the wide scale characterization of the thermal properties of buildings. *Journal of building performance simulation 11*(1), 65–83.

Shadram, F., S. Bhattacharjee, S. Lidelöw, J. Mukkavaara, and T. Olofsson (2020). Exploring the trade-off in life cycle energy of building retrofit through optimization. *Applied energy 269*, 115083–.

Shields, M. D. and J. Zhang (2016). The generalization of latin hypercube sampling. *Reliability engineering system safety 148*, 96–108.

Statistics Canada (2021). 2021 census of population. `https://www12.statcan.gc.ca/census-recensement/2021/dp-pd/prof/index.cfm?Lang=E`. [Online; accessed 25-February-2023].

Urge-Vorsatz, D., K. Petrichenko, M. Staniec, and J. Eom (2013). Energy use in buildings in a long-term perspective. *Current opinion in environmental sustainability 5*(2), 141–151.

USDOE (2023). Energyplus. `https://energyplus.net/`. [Online; accessed 25-February-2023].

Vaswani, A., N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin (2017). Attention is all you need.

Westermann, P. and R. Evins (2019). Surrogate modelling for sustainable building design – a review. *Energy and buildings 198*, 170–186.

Westermann, P., M. Welzel, and R. Evins (2020). Using a deep temporal convolutional network as a building energy surrogate model that spans multiple climate zones. *Applied energy 278*, 115563–.

Zhang, Y., Z. O'Neill, B. Dong, and G. Augenbroe (2015). Comparisons of inverse modeling approaches for predicting building energy performance. *Building and environment 86*, 177–190.

Østergård, T., R. L. Jensen, and S. E. Maagaard (2018). A comparison of six metamodeling techniques applied to building performance simulations. *Applied energy 211*, 89–103.