A SURROGATE MODELLING METHODOLOGY TO PREDICT ENERGY USE FOR MULTIPLE SINGLE-FAMILY CENTURY HOME ARCHETYPES IN TORONTO

By Cameron Rochon Lawrence

Mechanical Engineering (B.A.Sc.), Queen's University, 2017

A thesis

presented to Ryerson University

in partial fulfillment of the

requirements for the degree of

Master of Applied Science

in the program of Building Science

Toronto, Ontario, Canada, 2020

AUTHOR'S DECLARATION FOR ELECTRONIC SUBMISSION OF A THESIS

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners. I authorize Ryerson University to lend this thesis to other institutions or individuals for the purpose of scholarly research. I further authorize Ryerson University to reproduce this thesis by photocopying or by other means, in total or in part, at the request of other institutions or individuals for the purpose of scholarly research. I understand that my thesis may be made electronically available to the public.

A SURROGATE MODELLING METHODOLOGY TO PREDICT ENERGY USE FOR MULTIPLE SINGLE-FAMILY CENTURY HOME ARCHETYPES IN TORONTO

Master of Applied Science 2020, Cameron Rochon Lawrence

Building Science, Faculty of Engineering and Architectural Science, Ryerson University

Abstract

This research investigates the application of surrogate modelling to improve the energy performance of single-family homes. EnergyPlus was used to simulate 6000 energy models for four different semi-detached and detached century archetypes in Toronto, ON. Multivariate regression and a novel forward stepwise selection methodology were used to develop the surrogate models for each archetype. These models predicted energy use between 7.02%-7.54% error. A combined model that contained all four archetypes was developed to determine if a single model can replace multiple models. This model predicted annual energy use with 7.03% error and the number of samples required per archetype was reduced by a factor of 3-4. Elastic net regression was tested and found to be equally as effective as the proposed stepwise selection methodology. The findings of this research support the future application of surrogate modelling as a powerful tool to develop bottom-up archetype models for century homes in Toronto, ON.

Acknowledgements

Thank you to my supervisor Dr. Russell Richman for supporting me in my academic journey, and for giving me the encouragement I needed to commit to writing this thesis. Thank you to my committee members Dr. Mark Gorgolewski and Professor Jenn McArthur for providing valuable insight and sharing their expertise.

Thank you to past researchers who have worked on the Toronto Archetype Project who have helped provide a foundation for this work. Thank you to Erica Barnes for sparking my interest in surrogate modelling and encouraging me to pursue this research. Thank you to my family and friends. Thank you to my parents for sharing invaluable research expertise and architectural knowledge, and for always supporting me. I would also like to acknowledge and thank the homeowners who participated in the study, and the Pocket Community Association for their ongoing support and interest in this research.

Contents

List of Tables

List of Figures.

1 Introduction

This chapter will provide a brief introduction to archetype work and building energy surrogate modelling and define the importance of this work.

Problem definition

Residential buildings make up 17% of Canada's secondary emissions and 14% of greenhouse gas emissions [1]. Single-family residences have the highest energy use intensity of all housing types due to their high exposed surface area to volume ratio compared to high rise residential buildings [2]. Jermyn [3] found that among single-family houses, century houses, built before 1940 Have the highest energy intensity. 63% of residential energy use is attributed to heating, and only 1% of energy use is attributed space cooling. Most houses are heated with natural gas, which has a much higher carbon intensity than electrical sources [4], [5]. Reducing heating load can be achieved by improving the performance of the building envelope and the HVAC system [3].

Deep energy retrofits of existing single-family houses are a more effective method of reducing greenhouse gas emissions than a teardown and rebuild. The life cycle costs, carbon emissions, and waste associated with rebuilding a house mitigate operational savings offered by this approach, and retrofitting existing houses, was found to be a more sustainable solution [6].

Modelling homes individually is not feasible given the size of the housing stock. It is far more efficient to work with archetype models to represent subsets of the housing stock. Blaszak [7] and Jermyn [3], have helped to develop the Toronto Archetype Project in order to characterize

the Toronto housing stock. This work has focussed on developing bottom-up archetypes using real measurements and energy modelling software to determine optimal retrofit sequencing for common house types. In an analysis of 23 urban Toronto neighbourhoods, Jermyn [3] found 35% of units (59,135 out of 165,825) were single-family houses. Century houses also made up 60% of this group, corresponding to 25,227 houses. Wartime, 70s Ontario Building Code and modern archetypes made up the remaining fraction of houses [8]. This research will build on the work by Jermyn and Blaszak.

Jermyn [3] identified a need for more archetypes to define each vintage within the Toronto housing stock. In an urban neighbourhood study, the archetypal three-storey semi homes made up only 51% of the semis, while two-storey semis made up 46%. Century Semi and Detached houses made up the majority of single-family houses in urban neighbourhoods [3]. Developing baseline models to describe two-storey semis enables a substantially larger percentage of century homes in Toronto to be represented. A field study was conducted in a neighbourhood east of downtown Toronto called "The Pocket". The results of this field study were used to expand the existing three-storey detached and semi archetypes to include two-storey semi archetypes.

There are two approaches to archetype development, the bottom-up approach and the top-down approach, which uses historical data to build archetype models. This thesis will build on the current bottom-up modelling strategy, which uses building physics and measured data to assess the energy-saving potential from various retrofit or technology options. Optimization algorithms are often used to determine the least expensive path to meet an efficiency target. To run an optimization, a large number of building energy models need to be simulated to

represent all the combinations of possible retrofits for each building feature. Most researchers reported that simplifications and concessions in model accuracy are often necessary to address computational limitations in large scale optimizations [9], [10]. Recently surrogate modelling has been used as a way to substantially reduce computational times associated with optimization and energy model development.

Several researchers [9] have investigated surrogate modelling as a hybrid bottom-up building archetype modelling strategy that combines building physics with machine learning algorithms. Surrogate models use a supervised machine learning approach to approximate the output of a complex and computationally demanding process by estimating the relationship between a set of predictor variables and an output(target) [9], [11]. Building energy surrogate models are trained using a synthetic dataset of buildings that have been simulated with energy modelling software. The set of buildings is developed based on a range of building features that describe the design space. Each building that has been run represents a single training sample that a learning algorithm will use to develop a surrogate model. Surrogate models have been developed for a variety of building types and design goals.

Only a handful of studies have investigated using surrogate modelling for single-family dwellings, and most of these studies have been top-down models [12] or were used for energy labelling instead of calculating end-use.

Only a single study was identified that investigated the use of a single surrogate model to model multiple buildings [13]. The process of developing surrogate modelling is time-consuming and complicated. Currently, the individual building-specific approach to developing

surrogate models limits the potential application of surrogate modelling to solve larger-scale problems. The feasibility of using surrogate models for large scale retrofit optimization problems is contingent on the scope of a single model being expanded to be able to accurately predict energy use for more than one archetype and a more comprehensive subset of the housing stock.

Four distinct single-family home archetypes with different characteristics were explicitly targeted in this research. These dwellings are all defined as Century homes (pre-1940). The four archetypes that were targeted are shown below in Figure 1.

| Three-Storey Semi | Large Two-Storey Semi | Small Two-Storey Semi | Detached |

Figure 1 Four archetypes investigated in this research

This research aims to determine the feasibility of creating surrogate models of various subsets of the stock of single-family century houses in Toronto in order to address shortcomings in the current archetype development approach. The four archetypes studied in this research are compared horizontally, and this research will investigate whether multiple different archetypes can be described by a single surrogate model (i.e. different number of floors, different roof shapes, detached vs. semi). Each archetype will be modelled using the same ranges in order to allow a horizontal comparison between archetypes.

This research will also investigate what machine learning algorithms and methodologies are appropriate to develop surrogate models to predict energy usage for single-family houses to simplify future research.

Objectives

Each chapter in this research will be divided into phase 1 and phase 2. The overall goal of this research was to investigate the application of surrogate modelling to expand and add value to the existing research on bottom-up urban Toronto archetypes.

In phase 1, a field study was conducted in a targeted neighbourhood to determine what archetypes exist and whether the existing Century Detached and Three-Storey Century Semi archetypes proposed by Jermyn [3] should be expanded to include a Two-Storey Semi archetype. The development of the dataset used in phase 2 also occurred in this phase.

In phase 2, Surrogate models were developed to determine how effectively a multivariate regression-based surrogate model can be used to predict energy usage for increasingly comprehensive subsets of the Toronto housing stock.

- The first investigation was used to determine what accuracy is possible with surrogate models trained and tested on a single distinct archetype.
- A second investigation, a single model was trained on a dataset, including multiple archetypes and was tested on individual archetypes.
- A third investigation was used to determine what accuracy is possible for a set of models that were trained on three archetypes and tested on an independent test set containing the fourth archetype.

The goal for these investigations is to determine potential accuracy for models developed on single, and multiple archetypes, and estimate a baseline accuracy that can be expected when a surrogate model is used to predict energy on the entire Toronto century housing stock.

Two different feature selection methodologies for a set of highly collinear feature options will be investigated for use in linear regression models to inform further research and compare options. A model will also be built using an artificial neural network to determine the level of accuracy that is possible for a more flexible learning algorithm. The research questions are outlined below.

1. After establishing the range of single-family archetypes present in a targeted urban neighbourhood in Toronto, how can an existing set of century house archetypes be expanded?

2. What accuracy can be achieved for four surrogate models developed to predict energy use for four distinct single-family residential archetypes, how does this differ from the accuracy achieved when using a single surrogate model for all four?

3. What is the most effective learning algorithm and feature selection methodology to develop surrogate models for bottom-up archetypes of century homes in Toronto.

Structure

The chapters in this thesis are divided into two phases. Phase 1 consists of field research, and a continuation of work on the Toronto Archetype Project. Phase 2 investigates the application of surrogate models to develop the Toronto Archetype Project further. The background section will focus primarily on modelling techniques used in this thesis and will also inform the methodology used in the results section.

2 Background and literature review

This chapter will provide a background for Phase 1 and Phase 2 of this research. In Phase 1 of this chapter, some of the current archetype work will be explored, and some of the gaps in the existing research will be identified.

In Phase 2 of this chapter, some existing applications of surrogate modelling for archetype development will be presented. This section will focus primarily on the background for the methodologies used in this research and will lay out the current best practices in surrogate model development.

Phase 1: Archetype Background

2.1.1   Bottom-Up vs. Top-Down Archetype Development

There are two main approaches to archetype development, top-down and bottom-up model approaches. M. Kavgic et al. [14] conducted a review of bottom-up building stock models. He describes top-down modelling as models built on aggregated level historical data and macroeconomic trends instead of physics or component-based factors. These models are not suitable for making technological decisions. Bottom-up approaches generally combine disaggregated components and physics-based analysis or building components [14]. Bottom-up approaches can be further divided into statistical, physics-based and hybrid models. Physics-based models use measurable data and often use energy modelling software or mathematical relationships to calculate energy use. These models help determine the effect of technology or retrofit strategies on energy use and can, therefore, help guide policy. The downside of these models is that they require more technical information than other model types. Statistical

bottom-up models can be created using utility data and can offer insights into macro and socioeconomic relationships to energy use. These models do not offer very much insight into policy decisions and energy reduction measures. Hybrid models use combinations of physics-based and statistical bottom-up models [15]. Figure 2 shows the general process involved in creating bottom-up archetypes.

Figure 2 Bottom-up archetype development steps

Swan and Ugursal[16], and Reinhard and Davilla [17] also conducted thorough reviews of residential energy modelling strategies. According to Swan and Ugursal [16], an archetype model is defined by characteristics such as vintage, number of storeys, geometry, building envelope and HVAC type. A single archetype model is designed to characterize a subset of the building stock. Building performance simulation is often used to estimate energy use for an archetype model and estimate the effect of a retrofit option on a subset of the housing stock [16]. Reinhard and Davilla [17] identified that a weakness in the current bottom-up archetype approach is that the method used to characterize the building stock and develop average can rely on a lot of assumptions that are often based on insufficient real measured data. If the base model is not robust the energy predictions will not be reliable.

2.1.2 Toronto Archetype Project

This research will primarily focus on the recent work conducted by several researchers from Ryerson University. These papers have followed the basic methodology of defining characteristic baselines, developing a bottom-up model and optimizing a set of retrofit scenarios to improve energy efficiency. Figure 3 describes the timeline of work conducted on the Toronto Archetype Project, which was first developed by Blaszak and Richman [7].
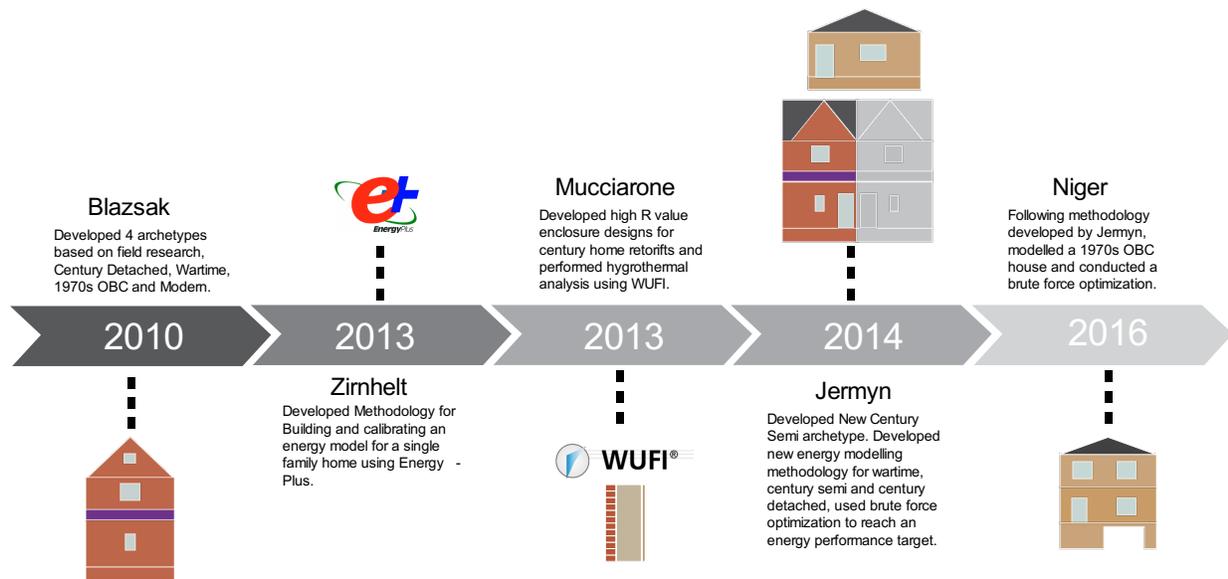
Figure 3 Timeline of archetype work in Toronto [3], [7], [18]–[20]

Blaszak divided the existing Toronto urban housing stock into four archetypal houses. These archetypes included a three-storey century detached house, a wartime bungalow, three-storey 1970s and a three-storey modern house. Base geometry and building performance parameters were taken from various datasets. The houses were modelled using Hot2000 [7] to determine baseline energy use. Several retrofit options were considered in order to achieve an energy use intensity below 100kwh/m^2. Blaszak's results suggested that building form and building type may have an impact on the optimal retrofit solution[7].

Independently to the Toronto Archetype Project, Zirnhelt [18] developed a methodology for building a calibrated EnergyPlus model of a single-family house in several Canadian climate zones, in order to model potential savings from passive solar design. Zirnhelt prepared a detailed methodology outlining the steps taken to produce a calibrated model.

Mucciarone [19] developed a set of building envelope retrofit designs for century houses in Toronto that delivered high R values and performed a detailed hygrothermal analysis to ensure that these wall designs would not cause any moisture issues.

Jermyn continued the work of Blaszak [3], by developing an additional three-storey semi archetype and adapting the model created by Zirnhelt [18] to create three calibrated EnergyPlus models to predict the baseline energy use for the century detached, and wartime houses proposed by Blaszak [7]. An additional three-storey century semi archetype was also developed. Jermyn [3], conducted a field study of 5 urban Toronto neighbourhoods In order to determine the baseline dimensions and feature inputs required to build these EnergyPlus models. A brute force optimization was conducted to determine a cost-effective route to an energy intensity of $70 kwh/m^2$. A brute force optimization was used because a full multi-objective optimization was not computationally feasible.

The three-storey century semi and three-storey century detached archetypes developed and the results of the field study conducted by Jermyn [3] were built upon in this research.

2.1.3 Bottom-Up Archetype Research in Canada

Several researchers have developed bottom-up archetype models for Canada. Swan et al. [21] developed *the Canadian Hybrid Residential End-Use Energy and GHG Emissions Model* (CHREM)

as a hybrid bottom-up model that is representative of the Canadian housing stock. Swan et al. [22] developed a nationally representative set of 17,000 real single-family homes referred to as *The Canadian Single-Detached and Double/Row Housing Database* (CSDDRD). The CSDDRD is a subset of the EnerGuide for Houses Database [23] which was developed by NRCAN [24] based on the data recorded during a nationwide energy audit program. Swan et al. [21] simulated each house contained in the CSDDRD using ESP-r [25] as the building simulation engine. Artificial neural networks were used to assign appliance and district hot water loads to each house based on specific profiles. Retrofit options were evaluated by re-simulating all the relevant CSDDRD models and determining the resulting energy and resulting and green-house gas savings using carbon emission factors for each end-use energy component. The CSDDRD models are representative sample of the entire national stock, so the results or the CHREM simulation can be extrapolated to represent the entire Canadian housing stock. [21]

Wills [26] revised the CHREM using an updated residential appliance and lighting model. Wills also included new envelope retrofit options and HVAC options such as ground source heat pumps in the CHREM model. This new model was used to perform a comprehensive analysis to determine the feasibility of retrofitting entire communities to net-zero using a variety of technologies. Community scale retrofits such as solar thermal, district heating and thermal energy storage were considered. In order to reach the net-zero target, a single-objective optimization was performed A multi-objective optimization was not feasible given the computational requirements [26].

In 2017 BC Housing and Integral conducted a large-scale optimization on small medium and large single-family houses in British Columbia. This report investigated the cost of achieving

various levels of a new building code based on predetermined code requirements that are required for each step on the path to net-zero. Hot2000 was used to reduce computational time compared to dynamic modelling. A built-in optimization tool, "Housing Technology Assessment Platform," was used to perform the analysis. Eleven building features were included in the optimization, and between 4 to 8 options were included for each feature. 54 million simulations were required to model a house with every combination of building features included in the design space. Running 54 million simulations was not feasible, so instead, a sample of 60,000-240,000 houses were simulated for each house.[27]

2.1.4   Limitations of the Current Bottom-Up Archetype Approach

Some common gaps in the current research on single-family archetypes are identified in this section. An archetype model is often based on insufficient real data and very general assumptions, which make the results unreliable for a large sample of the housing stock [3], [17]. Additionally, the current approach of using computationally expensive building simulation engines to build bottom-up models limits the potential application and analysis opportunities [21], [26]–[28].

Simplifications to the CHREM created by Swan et al. [21] were necessary due to inadequate data in the CSDDRD. Complex house geometries were reduced to a basic rectangular form in order to use the data provided by the CSDRD, this will likely affect the accuracy of the results. A sensitivity analysis or optimization would also not be feasible with this model. To test a single retrofit option takes approximately 20 hours on average using two computers with four-core processors that had to be purchased for the project [21]. Using a single processor this would take approximately 300 hours [29].

Using CHREM as a base model, Wills [26] was forced to perform a simplistic optimization using only a single objective due to computational limitations. A full multi-objective optimization that included cost would be optimal when considering the actual implementation of retrofits on the housing stock.

Jermyn [3] identified that simplifications often had to be made to energy models or optimization algorithms to enable optimizations. Instead of a full optimization, Jermyn used an iterative stepwise approach (brute force) that will not necessarily find the optimal solution. In the study conducted by BC Housing [27], only a small percentage of potential solutions could be tested due to the massive number of simulations required to model every combination. Even this small subset of potential combinations took several days to run for each archetype. Another limitation of the research conducted by BC Housing is that only one building form was analyzed to generalize the performance for many house shapes [27].

Jermyn [3] suggested that building an archetype model based on field measurements was challenging and potentially inaccurate. This potential limitation was also noted by Reinhart and Davila [17]. A single baseline model would not necessarily be representative of a large subset of the housing stock. Determining average house characteristics is challenging due to a small non-random sample, and variations in building form [3].

Surrogate modelling is investigated in the next section as a way to address the limitations of the current archetype approach.

Phase 2. Surrogate Modelling

This section will provide background for the subsequent chapters in this research. Three recent (2018-2019) reviews of the state of the current research on building energy surrogate models were primarily used to guide the research in this section. Barnes [30] reviewed 22 papers in her review, Westermann and Evins[9] reviewed 57 papers, Seyedzadeh [12] reviewed 40 papers. These papers are often cited in this section and should be accessed directly if a reader desires a more comprehensive background review.

2.1.5   What is Surrogate Modelling

Surrogate modelling is a term used to describe the use of supervised learning algorithms in order to approximate the output of a computationally expensive simulation. Surrogate modelling can be especially helpful for optimization tasks in which computational demand can prevent full optimization analysis from being conducted. [11]
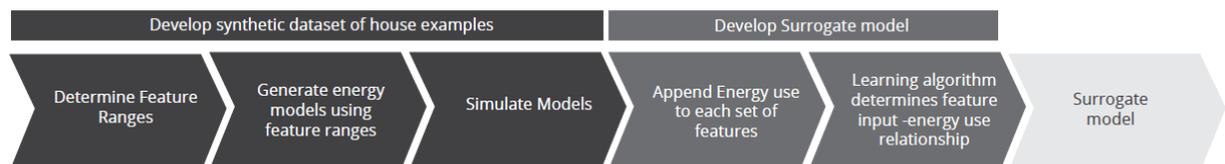
Figure 4 Surrogate modelling process.

It is important to differentiate the machine learning approach from the purely statistical approach. Statistical approaches often try to identify relationships and test hypotheses. The machine learning approach involves optimizing prediction accuracy using specific training and independent testing sets. "Statistics draws population inferences from a sample, and machine learning finds generalizable predictive patterns," As stated by Bzdok et al. [31].

14

Supervised learning refers to an approach in which the inputs are known and labelled. In unsupervised learning, the algorithm determines a relationship without set inputs or labels [11].

Recently advancements have been made in the field of surrogate modelling for use in building energy simulation. Where energy models may take minutes to hours to run, surrogate models have been developed that can run in less than a second while still providing an accurate estimate of energy use. This substantial reduction in computation time enables multi-objective optimization and quick estimates of energy use during the design stage.

2.1.6   Model Intent

Surrogate models have been developed for various applications and various target variables. The most common output is total annual energy[9], but often heating demand or cooling demand will be considered separately [32]. Although only one output is possible with multivariate linear regression, some researchers have considered using dummy variables to represent multiple outputs, such as end-use or monthly energy demand [30]. In artificial neural networks, it is possible to have multiple target variables by having an output node for each target considered. Models can be used to predict continuous numerical outputs (regression) or to group inputs (classification) [12]. Neural networks can be prone to overfitting and knowledge of hyperparameters and structure is essential to achieve high accuracy [12]. More complex algorithms such as Artificial Neural Networks (ANN) are ideal if prediction accuracy is important, and interpretability is not required.

Surrogate models are often used to reduce computing time for complex optimization problems. Linear regression models can be used to perform uncertainty or sensitivity analysis if care is taken to address collinearity among input variables [33], [9].

Westermann and Evins [9] found that surrogate models have been shown to significantly reduce computational times compared to using full dynamic models to test every iteration. Magnier and Haghighat developed a TRNSYS model of a single-family house, they found that it would take approximately ten years to run a full multi-objective optimization based on all the feature options. The entire simulation took only three weeks when using a surrogate model built with an artificial neural network to run the optimization. Simulation time to build the dataset was included in this estimate, running the optimization on its own only took 7 minutes [33].

Another important potential application for surrogate modelling is in guiding policy decisions for large scale retrofit projects or building energy codes [27]. Bottom-up surrogate models can be used to compare technology options and determine what energy reduction is possible. Top-down models can be used for energy labelling and classification [15].

Westermann and Evins [9] identified a significant gap in the literature he reviewed. Of the 57 papers, only a single paper had attempted to model multiple buildings to predict energy for a category of building types. Ascione et al. [13] created a model to estimate energy usage for office buildings in Italy and using artificial neural networks achieved an R^2 between 0.96 to 0.995 and a relative error between 2 to 11%. Although the number of floors and aspect ratio was modified, a basic rectangular shape was used for all models. Ascione et al. [13] proposed

that being able to model multiple building types would substantially improve the functionality of surrogate modelling. No research to date has attempted to develop a single surrogate model for multiple single-family houses using a bottom-up approach.

2.1.7   Learning Algorithms

The most common types of learning algorithms are multivariate regressions, followed by artificial neural networks [9]. Barnes[30] found that 19% of papers used artificial neural networks and 31% used multivariate regressions. Seyedzadeh et al. [12] found that artificial neural networks are one of the most common methods used in building energy surrogate modelling. Some other common algorithms include support vector machines (SVM)., random forest (RF) and multiple adaptive splines (MARS). Often the model intent informed the decision to use a particular model, and multivariate regressions were often favoured when interpretability of coefficients and sensitivity analysis was desired [9]. More flexible models such as artificial neural networks were chosen if prediction accuracy was the only consideration. Models can also be divided by classification and regression models; regression models predict a continuous output, whereas classification models are used to classify whether a building fits into a specific group using a binary output [9].

Barnes[30] found that papers in which multiple learning algorithms were compared, hyperparameters selection was often not discussed or disclosed. Without tuning hyperparameters, the results of comparisons between learning algorithms are meaningless. Seyedzadeh et al. identified a need for a fair comparison between different machine learning

algorithms. The current research does not provide enough information on the structure of the model and whether hyperparameter tuning was used. If the proper structure is not determined, the performance of ANN models is not guaranteed [12].

Multivariate Regression

Multivariate regression is an extension of a simple linear equation and is commonly used in a range of applications, including statistical analysis. Instead of fitting a single input to an output (target variable) as in a simple univariate linear regression, in multivariate regression, the function is fit to a plane in multidimensional space by reducing the least-squares equation to a minimum. The basic multivariate regression equation is shown below.

$$y_1 = \beta_0 + \beta_1 x_1 + \cdots \beta_n x_n$$

Artificial Neural Network

Artificial Neural networks (ANN) are inspired by the architecture of neurons in the brain. The algorithm learns by iteratively re-adjusting weights in order to reduce the cost (error) function until a minimum value is achieved [34]. The basic structure of a feed-forward neural network can be seen in Figure 5.
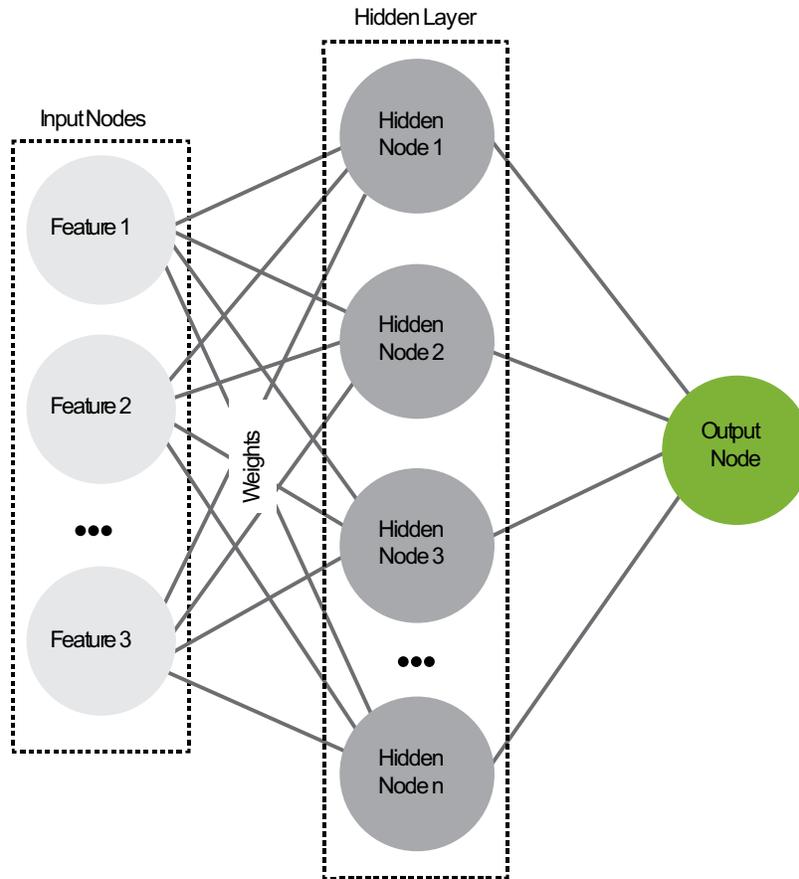
Figure 5 Neural Network Structure

Feed-forward neural networks (Multi-Layer Perceptron) are the most common types. A neural network consists of a set of input nodes (features) weights, one more hidden layers consisting of a number of hidden nodes, and one or more outputs nodes. The feedforward pass is shown in Figure 6.

$$f\left(b + \sum_{i=1}^{n} x_i w_i\right)$$
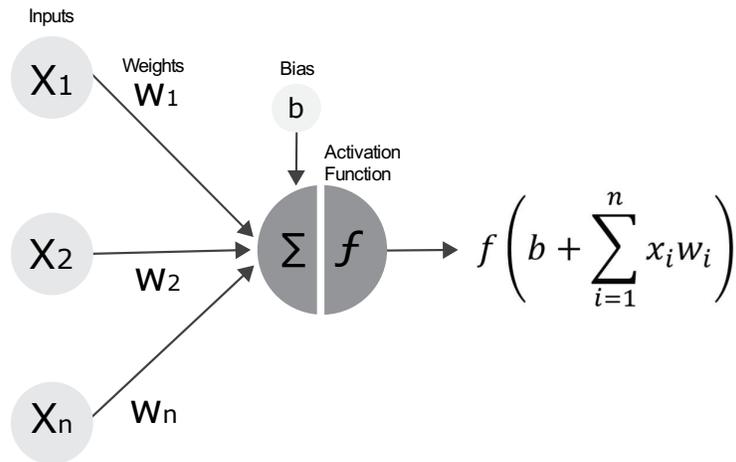
Figure 6 Feed forward algorithm.

Input features are multiplied by weights and then summed and transformed with an activation function before being fed to the next hidden layer or output node. [35],[36]. The result of a pass through the neural network is an estimate of the target value. This estimate is compared to the predicted value, and a cost(loss) function is used to compute the error. The neural network is trained through gradient descent in which the loss function is minimized. A feedforward neural network learns through a back-propagation algorithm. The slope of the error function with respect to each weight is calculated using partial derivatives and this value is used to update the original weight estimates through each backwards pass. The algorithm propagates back and forth and iteratively updates the weights until the loss function reaches a minimum. [35], [36]

A rectified linear function is often used as an activation function. This function returns zero if the input is less than one, or else a continuous value. Different activation functions are used for different purposes; they are often used to map a continuous numerical value to a value

between 0 and 1 for classification purposes. If an activation function is not used on the output node, the neural network can be used to predict a continuous numerical output similar to a regression model.

Many hyperparameters can be tuned in a neural network to improve performance. Without tuning hyperparameters in neural networks, high accuracy will not be likely [12]. Increasing the number of nodes in a hidden layer increases the complexity and flexibility of a model. This can lead to overfitting or underfitting. Some papers tuned the structure of the network, while others tested only one configuration. Very high prediction accuracy was possible in most cases with neural networks up to R^2 of 0.99 [12]. Changing the number of "epochs" determines the number of times the full dataset is passed through the algorithm. One epoch is one full pass through the neural network. "Batch size" refers to how many samples are fed through the network in one pass in order to update weights. Modifying these parameters is critical to ensure that the model converges and does not become overfit to the data. [35]

2.1.8   Evaluation Metrics

Many different performance metrics are reported between studies. There is not one single accepted metric in the literature. Each has applications where they fit better than others. The best approach is to look at more than one to get a sense of how well a model is fitting.

Barnes found that "mean absolute percent error" (MAPE), root mean squared error (RMSE) and "the coefficient of determination"(R^2) were the most common performance metrics found in the papers that were reviewed, and that these metrics were used in 15%, 26%, and 26% of papers reviewed [30]. Westermann and Evins Most of these studies reported more than one

error metric to allow comparison to multiple studies. RMSE is a useful tool for tuning hyperparameters or selecting between features within a model, although popular this metric is not useful for comparing between studies because it requires the scale of the target variable to be the same.

R^2 allows for a comparison of the goodness of fit between different models and studies. This metric is a measure of the amount of variance in the target variable that can be explained by the model. R^2 is a popular measure of the fit of the model to the data. Westermann and Evins [9] also found that R^2 was one of the most commonly reported performance metrics. R^2 will increase even with systemic error (all values are off by the same amount) or with a larger range of values being predicted, while average error may not improve. It will be easier to achieve high R^2 in buildings with larger energy use ranges, so it is helpful to report an error estimate as well. [37]

MAPE is a popular error metric because it measures the error as a percentage relative to the mean of the predicted result. MAPE can be used to compare between studies and has the advantage of being very simple and interpretable. This is important when considering a tool that may be used by non-practitioners who would not understand a metric such as RMSE or R^2

Acceptable error is difficult to quantify, given the limited number of studies that can be directly compared. If a comparison between studies is required, the papers would need to report the same metrics. In addition, the study would need to be looking at a similar-sized building in a similar climate zone. Ideally, the study would also need to report mean cross-validated performance metrics generated using the same learning algorithm. Westermann and Evins [9]

found that some regression models were able to achieve R^2 higher than 0.95 and found

prediction scores as low as 0.62, and in general, it is feasible to obtain R^2 above 0.95.

2.1.9 Preprocessing

The relationship between the target variable and the input variables can often not be captured

with a simple linear equation. Non-linear trends will introduce high bias to the model if a linear

equation is used. Higher flexibility models can be used to capture non-linear trends, but these

models often can often not be interpreted.

A model can be transformed in order to fit the data more accurately, while still maintaining the

underlying linear structure of a regression model. In some papers, the input or output features

are transformed in order to improve the regression model to the underlying data structure.[9]

Barnes [30] used a *Box-Cox* transformation to create a more normal relationship. The result

ended up being a logarithmic transformation of the input and output features. Some common

transformations include logarithmic, reciprocal and square root.

Standardizing the input features is an important step that should be performed before the

regression model is fit to ensure that the scale of the input features does not affect the fit of

the model. Standardizing scales the input features to have a mean of 0 and a variance of 1,

failing to standardize the results before running the regression will mean that the coefficients

and their relationship between each other can not be interpreted. In regularized regression,

this step is especially important because the scale of the variance and the input features will

determine how the model is fit and which features are dropped [38]. Standardizing is not

required for ordinary least squares regression, but failing to do so will mean that the coefficient

values cannot be compared to each other on the same scale [39]. Normalization is also sometimes used in regression. This means to scale inputs between 0 and 1. Although they mean different things, standardization and normalization are often used interchangeably. The mean of the input column is subtracted from each value and then divided by the standard deviation of the input column to standardize a dataset.

$$Z = \frac{x-\mu}{\sigma} \quad (1)$$

2.1.10 Model Validation

When creating a model, a dataset is split into a training and testing set. The model is fit on the training dataset and is validated using one or more performance metrics that measure how accurately this model can predict the target value in the testing set. A model must be chosen to minimize test error as opposed to training error. [38]

- Training set: data split used to generate the model by fitting the relationship between the input and output features.

- Validation set: Data split used to estimate error based on a hyperparameter value or addition of a feature. This split of data is used to make decisions on the structure of the model. Making models decisions using this set means that it is no longer an independent testing set, so validation accuracy is an optimistic estimate of the true error.

- Test set: Independent Data split used to report the accuracy of the model developed with the training+validation set. [37]

Training error only describes how well the model can predict based on data that it has already seen. A highly complex model such as a neural network may be able to fit a training set perfectly but will prove to be inaccurate when exposed to new data. Finding the right amount of complexity in order to predict accurately on the test set is referred to as the "bias-variance trade-off"[38].
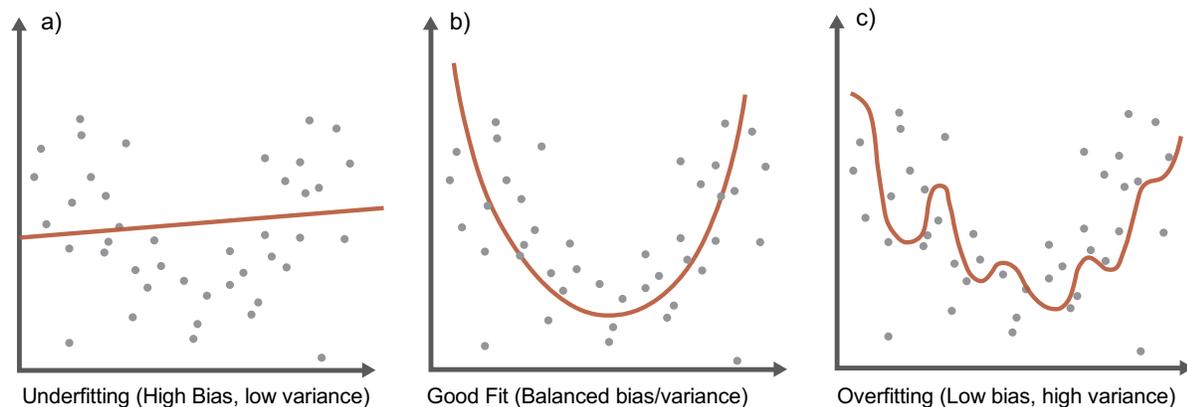
Figure 7 Bias variance trade-off, a) underfit model b) a good fit c) overfitting

Bias in a model refers to its complexity and how well it fits the training set. A low bias model is a complex model that minimizes training error. A low bias model may pick up on noise in the training data that does not follow the overall trend of the data; this example of overfitting can be seen in Figure 7. A high bias model is a simplistic model that has a higher training error and may miss some important trends in the data.

A low variance model is a more simplistic model in which the difference between the testing and training scores is minimal; this model will be able to generalize well to new data. In a high variance model, the training error is a poor estimate of testing error, and often this occurs

when the model has low bias and therefore fits the training data too closely. When exposed to unseen data, the trends picked up in the training data are no longer present, and the error will be unpredictable.

A dataset should be split into training, validation and test sets to allow model tuning without compromising the integrity of the accuracy estimate. The validation set is used to tune hyperparameters and select features, and the test set is used to report the accuracy. Hyperparameters and features should not be chosen on the same dataset with which the error is reported. Selecting the optimal set of parameters based on the test set is a form of training, meaning that the test set would no longer be independent.

Resampling techniques such as cross-validation are commonly used to determine the mean and variance of the model that is more representative of the entire dataset instead of just one split of the data [38]. Using only a single split will not provide an accurate estimate of model prediction ability, especially with smaller datasets.

Some common cross-validation techniques include bootstrapping and k-Fold cross-validation. In k-fold cross-validation, the dataset is sampled in k subsets; each subset is used as a validation set exactly once, and the remaining subsets are used to train the model. K-fold validation is less computationally demanding than bootstrapping and provides a reasonable estimate of model performance; five or ten splits are often used [38]. If feature selection and hyperparameter tuning is performed, a nested cross-validation approach should be used, different cross-validation loops are required to report the accuracy and to tune the model [40]. An Inner cross-validation step is performed on training+ validation sets to tune the model. an outer cross-

validation step is used to vary the testing set [40]. No studies were found to employ nested cross-validation when hyperparameter tuning or feature selection was performed. Barnes [30] found that only 20% of the papers she investigated used cross-validation techniques. Westermann and Evins [9] did not mention cross-validation in the comprehensive review of the literature they conducted.

Barnes found that the number of samples and the proportion of training to test/validation sets varied between studies, and some studies did not perform analysis to determine if the sample size used was adequate given the number of inputs [30]. Hygh et al. [32] used regression and energy plus to predict the performance of buildings in various climate zones, 20000 samples were simulated initially. A sample size analysis was performed, which revealed that only 1000 samples were sufficient to achieve the same level of accuracy for the building studied. Chen et al. [41] found that the RMSE and R^2 stabilized in their analysis past 600 samples and that the number of input features was more significant at smaller sample sizes. Barnes[30] found that MAPE started to increase significantly only when the number of samples was dropped below 1000.

2.1.11 Sampling Methods

When building a dataset, several common sampling techniques can be used. These techniques can be categorized as being either static or dynamic and can have a substantial effect on the accuracy of the model [42], [30]. Latin hypercube sampling creates a matrix of random space-filling values between 0 and 1. A comprehensive analysis comparing performance with several sampling techniques has not been conducted. Westermann and Evins [9] found that the

majority of papers used Latin hypercube sampling, 81% used a static sampling technique

compared to 19% that used an adaptive sampling technique.

2.1.12 Preprocessing

Features that cannot be described with continuous numerical ranges, or in which a linear

relationship to the target variable is not expected can be modelled using categorical variables

instead.

There are several approaches to dealing with categorical variables, including "one-hot

encoding" (dummy encoding) and "label encoding". [43] [38]. When one-hot encoding is used,

each option in a category feature is re-encoded into "dummy" binary features. 1 signals that an

option is selected, a 0 signals that the option is not selected. An example of one-hot encoding is

shown below in Table 1, the table on the left is the categorical variable pre dummy encoding,

and the encoded version is shown on the right.

Table 1 One hot encoding Example for Light Type

	Light Type			Fluorescent	Incandescent	LED
House 1	Fluorescent		House 1	1	0	0
House 2	Incandescent		House 2	0	1	0
House 3	Fluorescent	→	House 3	1	0	0
House 4	LED	→	House 4	0	0	1
House 5	LED		House 5	0	0	1
House 6	Incandescent		House 6	0	1	0
House 7	Fluorescent		House 7	1	0	0

When dummy variables are used, it is important to drop one of the features when used in

unregularized regression. When using regularization steps such as lasso or ridge, all n options

should be left in.[39]

Label encoding is another option for dealing with categorical variables. Label encoding replaces each option in a category with continuous numerical values, e.g. led=1, fluorescent =2, incandescent =3. The problem with this approach is that if the designer places a larger value to denote a certain option, the model will interpret this as 3>2>1 when this may not be the real relationship [38].

2.1.13 Collinearity

When two or more features are highly correlated to each other, the result is collinearity. Collinearity does not affect accuracy, but the presence of collinearity means that there is more than one optimal solution to the linear equation. Collinearity will result in the inflation of the variance of the regression coefficients, which can make it challenging to interpret the coefficient values. Severe collinearity can even result in the signs of the estimated coefficients being reversed [39]. Additionally, if two collinear features are included in a model, if the model is trained on a dataset with a collinear relationship, it can not be expected to predict well on a dataset without the same collinearity structure [44].

A common method to address collinearity is pre-filtering to remove collinear features before training. Pearson correlation calculates the univariate pairwise relationships between each feature, a value between -1 to 1 is expected. A value above .7 is considered a cause for concern. Variance inflation factor (VIF) is calculated by regressing each feature on all other input features in the model. The equation is shown below. A VIF value above 1 indicates some multicollinearity if the R^2 value is close to 1, the VIF value will approach infinite. The VIF value represents the amount of inflation of the variance of coefficient estimates that can be expected for each feature [44].

$$VIF_k = \frac{1}{1 - R^2{}_k}$$

The Pearson correlation on its own will often not be able to diagnose a multicollinearity problem. This occurs when a variable can be described by a combination of other features in the model. For example, volume may not have a strong correlation to any single feature (Pearson correlation greater than 0.7) but when regressed on width, depth and height, the VIF value can be almost infinite because most of the variability in volume can be explained by the other input features [44]. A filtering method such as VIF that addresses multivariate relationships is a much more robust tool for addressing collinearity than Pearson correlation that only addresses univariate relationships. A comprehensive linear algebra-based explanation of the cause of collinearity in regression models can be found in [45], this paper also confirms that Pearson correlation is not always effective on its own at addressing multicollinearity.

The threshold in which VIF is considered a problem is somewhat arbitrary and depends on the purpose of the model. A value between 5 and 10 is often used as a threshold to determine what features can be included, 10 is considered the point in which there is problematic collinearity [44], [46], [47], [39].

Regularization to Address Collinearity

Regularization is an embedded technique in which a penalty is added to the least-squares estimate. Ridge, elastic net and lasso are the most common regularization techniques. Ridge regression adds a squared value to the estimate, which reduces coefficients towards 0 but not to 0. Lasso regression adds an absolute penalty, which can shrink coefficients all the way to 0 given a large enough choice of tuning parameter (l). Elastic net is a combination of ridge and

lasso that averages the two effects [38]. Adding a penalty to the estimate introduces bias, which pulls the coefficients towards 0. Adding bias will counter the large effect that an outlying data point can have on the least-squares estimate and the resulting coefficients. When features are almost fully collinear, there can be an infinite number of potential solutions planes and the variance will be very large.[48] The equations for ridge and lasso are shown in the equations below.

Regularization is a popular way to remove collinearity in a set of features and has been shown to produce a more interpretable model than complex tools such as Principal component analysis in removing collinearity [44],[39].

2.1.14 Feature selection

Often the base inputs used in a surrogate model do not predict the target variable well; in addition, often, the base inputs include features that do not have significance in predicting the target variable. Features may also be collinear, and in order to improve interpretability, some features could be dropped.

The goal of feature selection is to include only useful predictors and remove irrelevant or redundant features. Feature selection methodologies can be categorized into filter, wrapper and embedded selection. Filter selection uses univariate rankings to determine the optimal features prior to training models. Wrapper selection uses trained models with subsets of features and compares performance metrics to select features. Embedded selection selects features with the learning algorithm Forward stepwise feature selection is a technique in which

features are tested iteratively by adding features from a list to the model one at a time and

keeping only the feature that improves the accuracy the most. [49]

Elastic net for Feature Selection

Elastic Net and Lasso regression are popular tools to reduce collinearity and improve

generalization ability. Regularized regression can also be used to select features. Lasso will

perform continuous shrinkage of the coefficients and feature selection simultaneously. If two

highly collinear features are included in a model, lasso will tend to drop one feature at random.

Elastic net was developed to address the shortcomings of lasso regression. Elastic net tends to

group collinear features and average the effect on the target variable instead of dropping one

feature randomly [50]. The equations for ridge, lasso and elastic net are shown below

$$RSS + \lambda \sum_{j=1}^{p} \beta_j^2 \quad \text{(ridge)}$$

$$RSS + \lambda \sum_{j=1}^{p} |\beta_j| \quad \text{(lasso)}$$

$$RSS + \lambda \left(\frac{1-\alpha}{2} \sum_{j=1}^{p} \beta_j^2 + \alpha \sum_{j=1}^{p} |\beta_j| \right) \quad \text{(elastic net)}$$

In the equation for elastic net, a is known as the L1 penalty coefficient. The L2 penalty

coefficient is 1-a. When a is 0, the properties of elastic net are the same as ridge, and

coefficients will be shrunk but not dropped all the way to 0. If a=1 elastic net functions as lasso

regression, features will be shrunk all the way to 0.

Forward Stepwise Feature Selection

Currently, forward stepwise selection has some disadvantages. This tool will not always arrive

at the most accurate model. In order to determine the true optimal combination of features,

every possible combination of features would need to be considered; this would be computationally limiting. Instead, forward stepwise tests features iteratively by adding them one at a time. Forward stepwise selection also traditionally does not address collinearity when adding new features to the model that may be linear combinations of other features in the model [39].

Using stepwise selection Hygh et al. [32] improved the R^2 of their regression model from between 0.859 to 0.977 to greater than 0.96, by increasing the number of features in the model from 27 to between 37-45. It is unclear whether features were selected independently from the test set, there was no mention of a validation set in the study, which would optimistically skew the reported accuracy. Significant collinearity issues due to the addition of features that were combinations of the original input features would be expected after the stepwise selection. There was no mention of this issue, which would affect the interpretability of the standardized regression coefficients used in a subsequent sensitivity analysis.

Only one paper was found which integrated a multicollinearity threshold into the feature selection process. Vu et al. proposed a backwards stepwise methodology to select the optimal features for a regression model and address multicollinearity. In this study, climatic features were selected to estimate electricity use. All features were included in the model initially, and features were removed iteratively in order of highest collinearity, VIF was tested again after each feature was removed a backwards stepwise regression was then performed on the features that remained when the VIF for each feature fell below the threshold of 5. The backwards selection was then performed using hypothesis testing with a two-tailed p-value.

This methodology addresses multicollinearity effectively by running VIF calculations iteratively. [45]

A flaw in this approach is that features with high collinearity, in this case, would be removed first. These features are likely collinear because they capture the effects of the largest number of coefficients and, therefore, would have the most explanatory power. Using the methodology outlined above, features with the most substantial relationship to the target variable may be removed before the actual backwards stepwise selection begins.

Barnes [30] used a novel combination of stepwise and lasso regression to select features. Forward stepwise selection was used to test linear combinations of the 71 original input features. Features were added in order of decreasing Pearson correlation. Features that improved the model were included, and the features used to generate these combined features were removed, in an attempt to reduce collinearity. This methodology reduced the number of features from 71 to 45 and improved the validation R^2 from 0.9643 to 0.9710. Barnes [30] performed a lasso regression to remove any redundant features remaining in the model and further reduce the complexity of the final model. Twenty additional features were removed without significantly affecting accuracy. A univariate correlation analysis was performed on the remaining features and found Pearson correlations up to 0.9 in the remaining features. The lasso regularization step would reduce collinearity, so a univariate analysis of the unregularized coefficients would not necessarily diagnose collinearity in this context. The extent to which

collinearity is reduced through elastic net, lasso or ridge regression has not been studied in the

context of building science surrogate modelling.

3   Methodology

Phase 1:

The general steps taken for phase 1 is outlined below.

Steps:

1. Prepare and obtain Research Ethics Board approval.

2. Identify common century semi house types and geometry features in a defined neighbourhood in Toronto neighbourhood.

3. Measure accessible features of selected houses (e.g.: width, height, window to wall ratio, roof shapes, number of floors.)

4. Based on field measurements, model two new archetypes in EnergyPlus following the methodology proposed by Jermyn [3].

5. Using Data from Jermyn [3], field measurements and background research determine the range of feature values that will reasonably capture the population of century house configurations in Toronto (from baseline existing performance to passive house performance targets).

An urban Toronto neighbourhood ("The Pocket") east of downtown Toronto was targeted to develop two-storey semi archetypes. Jermyn Identified that finding willing homeowners to be included in the study as challenging. The community group in this targeted neighbourhood was already actively engaged in researching potential strategies to reduce the carbon footprint of their neighbourhood through targeted retrofitting strategies. The Pocket neighbourhood was designated by the City of Toronto as one of two "Community-Led projects on house

retrofitting" [51]. The presence of willing participants and a high density of two-storey semi archetypes made this an ideal neighbourhood to target for the field study.

Two-storey semi houses made up 46% of single-family century house archetypes in surveyed neighbourhoods; for this reason, they were targeted by this study [3]. Google street view and observational study in the neighbourhood were used to identify house shapes that frequently appeared (more than ten times). These houses were recorded and grouped. Google Earth [52] has a 3d view that can be used to measure dimensions in 3 dimensions in order to get an estimate of the building geometry and define sub-groups. If homeowners agreed to participate in the study, a visit was set up to perform data collection. Houses were measured to determine whether preliminary groupings were accurate and determine the range of characteristics as outlined in Table 2. Occupants were contacted through door to door canvassing, and an email announcement from the pocket neighbourhood committee. A series of forms were created in order to record information, a set of standard dimensions were developed for each house in order to simplify the measurement of exterior dimensions and allow average baseline dimensions to be calculated once the data collection was finished. Only accessible building parameters were measured, roof dimensions estimated using laser measures from ground level and interior dimensions. Some houses had access to the attic, which allowed non-destructive exploration of the building envelope. The methodology used to determine the baselines was based on an iterative approach of refining groups of houses into more specific sub-groups based on measurements of 18 houses and observational study.

A description of the data collected is outlined in Table 2.

Table 2 General overview of data that was collected in the field study

Geometry
Building footprint
Storey height
Exterior Dimensions
Glazing and doors dimensions and locations
Floor plan and Dimensions
Shading devices and overhangs
Envelope
Materials and material properties
Window constructions including frame, divider, reveal, and sill
Basement materials and material properties
Wall and floor thickness
Depth of wall below grade
Height of wall above grade
Internal Gains
Types of major appliances
Occupancy schedules more detailed (appliances and setpoints)
HVAC
Thermostat location
Type of heating and cooling
Furnace air conditioner efficiency and sizing
Utility Data

Standard dimensions that were recorded for each house type were organized in an excel spreadsheet. The data was organized and mean values were calculated for each dimension. Developing a set of standard dimensions to measure between houses was a critical step that simplified the process.

A survey was developed to determine occupancy scheduling, and the results of this survey are reported in the appendix. A survey was also used to determine what were considered the biggest obstacles in conducting energy efficiency renovations, potential budget ranges, energy

savings goals, and what building characteristics, if any, should be preserved if retrofits were considered. The survey and the door to door letter are included in the appendix.

Phase 2: Surrogate Modelling

The general steps taken in phase 2 are outlined below.

1. Develop synthetic dataset of 6000 total EnergyPlus models (*.idfs*)

2. Run multivariate regression with 30 base input features.

3. Develop 37 additional geometry features (volume, total surface area, house type) to test if better accuracy is possible with combinations of the base input features.

4. Develop automated feature selection methodology with collinearity threshold to determine optimal set of features for high accuracy and interpretability.

5. Use this methodology to develop surrogate models for each archetype and develop a single surrogate model for all four archetypes combined in to one model to answer the second research question.

6. Answer the third research question by testing two alternative methodologies for the combined model. Test elastic net regression as an alternate learning algorithm/feature selection methodology. Test artificial neural networks as an alternate learning algorithm to achieve higher accuracy.

In this research, the application of surrogate modelling was further developed and served to fill in gaps and add value to the existing body of work. The main objective of phase 2 was to investigate the use of a single surrogate model to predict energy for multiple archetypes and compare methodologies for feature selection and model development.

A novel approach to forward stepwise feature selection that addresses collinearity was developed. Using the stepwise feature selection methodology, a series of models were developed to investigate how surrogate models could be used to predict the energy use for a category of archetypes. An elastic net regression approach that is simpler to implement was compared to the stepwise approach for the combined model to see if one feature selection methodology offered clear advantages over the other.

A preliminary analysis was carried out to investigate the application of artificial neural networks and compare the pros and cons of multivariate linear regression compared to a more complex learning algorithm based on accuracy and research objectives.

The workflow for this thesis is laid out in Table 3. The framework of this table is based on a helpful best-practice strategy proposed by Barnes [30] in which important model decisions and strategies are clearly reported to allow another researcher to quickly compare their model and assess the validity of their findings.

Table 3 Surrogate model development methodology

Dataset Development	Model Intent	-Early stage design tool -Tool for optimization -Sensitivity analysis
	Target Variable	-Total Annual Energy
	Building Archetype	-Single-family Residential (4 archetypes, Including 2 Storey Semi 3 Storey Semi and 3 Storey Detached
	Location+Climate	-2016 Toronto City Centre, Ontario, Canada – CWEC weather file [53]
	Energy Simulation Software	-EnergyPlus v8.9 [54]
	Statistical Analysis And modelling tool	-Python v 3.7.4 [55]
	Base model	-Based on existing calibrated model, updated with IECC prototype models
	Features+Ranges	-see Table 4, Table 7, Table 9,

	Sampling Plan	-Latin Hypercube
Data Preprocessing	Train/Validation/Test Split	-kfold cross-validation (20% Test, 8% Validation, 72% Training), 10-fold inner loop, 5-fold outer loop - Used: sklearn.model_selection.KFold [56]
	Feature Engineering	-Input feature Standardization, mean of 0 standard deviation of 1 -Used: sklearn.preprocessing.StandardScaler() [56] -Logarithmic Transformation of input and target variables
Model Development	Feature Selection	-Stepwise Selection, with VIF threshold of 5 to prevent collinearity -Elastic Net Regression -Used: sklearn.linear_model.ElasticNet [56]
	Learning Algorithms	-Multivariate Regression Used: sklearn. linear_model.Linear_Regression() [56] -Elastic Net regression -Artificial Neural Network using: Kera's and TensorFlow[57]
	Error Metrics	-MAPE - R^2 -RMSE

The three major components of surrogate model development include dataset development, data preprocessing and model development. Without a dataset, a surrogate model can not be created. The dataset provides the training examples from which the machine learning algorithm learns. The surrogate model Is only as robust as the base model used to develop the dataset so ensuring this model was built on correct assumptions was critical.

Data preprocessing is an essential step in which the inputs and outputs are converted into a form that can be interpreted by the machine learning algorithm. Non-linearities in the results are addressed in this step by transforming the input feature or target variables if required.

The training, validation and testing of the surrogate model occurs in the model development step.

3.1.1 Dataset Development

Creating the dataset for this thesis followed the methodology shown in Figure 8.

Field Data Collection & Research

Deveope baseline archetype and energy plus model

Determine input features and ranges

MATLAB

Generates Latin hypercube matrix of random values evenly distributed from 0-1

Multiply each column in the latin hypercube matrix by its associated feature range.

(min+ (random value*range)

Create matrix of input features

Export input matrix to python as csv

python

set of 6000 .idfs

= 40

eppy.modeleditor

Reads CSV file containing randomized feature ranges for houses

Write script to automatically update energy plus .idfs using csv.

Generate 4x1500 .idf files representing the range of potential geometry and performance parameters

.idfs

Run 6000 .idfs

results

Extract energy use data and append to input matrix to create csv containing an array of input features for 6000 house samples with annual energy as the target variable

scikit learn

Import Input/target csv into python Develop surrogate model using scikit learn

Figure 8 Dataset development methodology

In order to create a dataset to train a machine learning algorithm, a set of 6000 EnergyPlus models *(.idfs)* were generated. Based on the findings of the literature review 500-1000 training examples are required per archetype model. To be safe 1500 *.idfs* were simulated for reach archetype . These energy models were generated based on a set of randomized input ranges that describe the design space that the surrogate model will cover. The only data that is required from each energy model is the set of input feature values and the total annual energy use. The total annual energy is the target variable. The surrogate model will be trained to use a set of input feature values to predict the target variable.

The baseline energy models were developed based on the input values used by Jermyn [3]. These inputs are listed below in Table 4. These inputs are the characteristic values that describe the baseline archetypes, and these inputs were modified in phase 2 to produce the surrogate models.

Table 4 Baseline Energy model inputs adapted from Jermyn [3]

	Semi Archetypes	Detached Archetype	Reference:
Roof Assembly			
Structure	Double Wythe Birck	Double Wythe Brick	Blaszak and Richman [7]
Cladding	Brick	Brick	Jermyn and Richman [3]
Insulation			
Ceiling (RSI)	2.64	2.64	
Walls (RSI)	1.01	1.01	Blaszak and Richman [7]
Foundation (RSI)	0.55	0.55	
Ceiling Assembly			
	Shingle	Shingle	
	OSB	OSB	
	100mm Fibreglass	100mm Fibreglass	Blaszak and Richman [7]
	Lath and Plaster	Lath and Plaster	
Wall Assembly			
	Double Wythe Brick	Double Wythe Brick	
	20mm Air Space	20mm Air Space	
	25mm Fibreglass	25mm Fibreglass	Blaszak and Richman [7]
	Lath and Plaster	Lath and Plaster	
Foundation Assembly			
	Quadruple wythe Brick	Quadruple wythe Brick	Blaszak and Richman [7]
Air Leakage (ACH 50pa)	11.7	10.4	EcoEnergy Database [58]
HVAC System			
Type	Forced Air Gas	Forced Air Gas	
Control Location	Dining Room	Dining Room	Jermyn and Richman [3]
Vent Flow Rate (m³/s)	1.3	1.3	
Internal Gains			
Lighting(w/m²)	1.9 and 3.07	1.9 and 3.07	Zirnhelt [18]
Appliances	Kitchen/Hot Water	Kitchen/Hot Water	Jermyn and Richman [3]
Occupancy	4 people	4 people	
Framing Factor			
Exterior Wall (%)	31	31	Qasass et al. [59]
Interior Wall (%)	15	15	Zirnhelt [18]
Floor (%)	11.7	11.7	Qasass et al. [59]
Roof (%)	8.6	8.6	

For the surrogate model, 17 numerical features and four categorical features representing 13 total options were chosen based on the inputs used by Jermyn [3]. The ranges for these features were designed to capture the existing observed baseline condition in the field all the way to a high-performance level (i.e. 'passive house ready' as defined by [3]). The sources for each range can be seen in Table 5, Table 6 and Table 7. Most of the ranges were selected based on past research on Toronto archetypes [3], [7]. The maximum and minimum measurements found across both the current and 2013 field studies were used to create the range of geometry values. Each geometry range was expanded by ±10%. The range of values was expanded to create a more inclusive range representative of the population of houses in Toronto, given the limited sample size available.

Some features are best represented by options instead of numerical ranges or are unlikely to exhibit a linear monotonic relationship between input and target variables. In these cases, categorical features can be used instead of continuous numeric features. Orientation is an example of a feature that will likely not have a linear relationship to energy use throughout a range from 0 to 360.

Regression models can only interpret numeric inputs, so *One-hot encoding* was used to convert categorical features into individual columns with binary values for each option. In the multivariate regression 1 option was dropped to eliminate collinearity, the option that is droppedd is considered the baseline with which the other options are compared. All options are included in elastic net regression.

Categorical features are listed in Table 5.

Table 5 Categorical Features

Category	Options
Light Type	Incandescent, Fluorescent, LED
HRV Efficiency	0,60,85,90
Orientation Azimuth	343/0, 73/90, 153/180, 253/270)
Neighbouring House Distance	1m , 10m

EnergyPlus v8.9 [54] was used as the building simulation engine to generate samples. The

energy plus inputs for each light type are shown in Table 6. The radiant and convective fractions

of heat gain were adjusted based on the light type used. These inputs are necessary because

the effect of the choice on heating and cooling demand should be considered in addition to the

effect on electricity demand. It is not clear if these values were changed in other studies.

Table 6 Lighting Inputs

	Incandescent	Fluorescent	LED	Source:
Light Intensity (up/downstairs)	1.94/3.07 w/m^2	0.605/0.958 w/m^2	0.314/0.498 w/m^2	
Fraction Radiant	0.73	0.37	0	[60]
Fraction Convective	0.8	0.21	0.25	

The ventilation rate was set based on the ASHRAE 62.2 Standard for each house [61], shown in

the equation below, where Q total is the total mechanical ventilation, Q_{fan} is the amount of

required mechanical ventilation in L/s after the infiltration credit is calculated based on the air

changes of the house. The infiltration credit Q_{inf} can not be more than 2/3 of Q_{tot}. A_{ext} is the

ratio of exterior surface area to the shared area (attached garage, semi-detached etc..). A value

of 1 is used for A_{ext} in detached houses.

$$Q_{tot} = 0.15 * Floor\ Area + 3.5 * (Number\ of\ Bedrooms + 1)$$

$$Q_{fan} = Q_{tot} - (Q_{inf} * A_{ext})$$

Houses with lower ACH50 values will have a higher requirement for mechanical ventilation, so the model must capture this effect. When running regressions, this meant that ventilation and various features such as volume were very highly correlated. The mechanical ventilation rate was normalized by the building footprint to eliminate collinearity.

The ranges of input features for numerical features are shown in Table 7.

Table 7 Base energy plus inputs

Final Range					
Feature:	Unit	Min	Max	Max-Min	Source:
Depth	m	8.035	17.265	9.23	
Width	m	3.6365	7.0335	3.40	
Above Grade Foundation Height	m	0.6	1.3425	0.74	
Below Grade Foundation Height	m	0.8	1.93	`	Field Study and [3]
Ceiling Height Average	m	2.1	3.255	1.16	
WWR Left/Right Side	ratio	0.003	0.138	0.13	
WWR Front	ratio	0.008	0.265	0.26	
WWR Back	ratio	0.008	0.265	0.26	
Window U-Factor	w/m2k	.125*	1*	.875	Field Study and [62], [3]
SHGC	none	0.200	0.700	0.50	
R-Value Slab	m²K/W	0.208	6	5.79	
R-Value Basement	m²K/W	0.706	10	9.29	Field Study and [3], [19]
R-Value Wall	m²K/W	0.652	10	9.35	
R-Value Roof	m²K/W	0.479	14	13.52	
Air Changes per Hour	ACH	1	21.77	20.77	[63]
Furnace Coil Efficiency	%	0.692	0.98	0.29	Field Study and [3]
COP Air Conditioning	COP	2.05	5	2.95	

*The U-factor range should be 0.7 to 2.7 the range used is unrealistically low due to an error in unit conversion.

Most of these values are taken from the current field study and the field study conducted by

Jermyn for a total of 35 houses from which to build the ranges. The same ranges were used for

each house type. Using the same ranges allowed for comparison between models. The data

collected for each individual archetype was not considered sufficient to estimate the true range

accurately, so the range developed from the much larger set of all archetypes measured was

used for each archetype individually. The maximum air changes per hour value represents the

95th percentile of the values found in the EcoEnergy Database for single-family homes built pre

1940 [58].

In order to develop the dataset, a matrix of random values was required to create a

corresponding matrix of input feature values. MATLAB's *lhsdesign* [64] sampling was used to

generate a *latin hypercube matrix* of 1500x22 randomized values covering a range from 0 to 1.

Each column of the *latin hypercube matrix* was used to generate a set of random feature values

for a single feature. Each element in the column was multiplied by the range(max-min) and

then added to the minimum value. Each row of the resulting matrix represents a single house

example based on a set of randomized building features. This step was repeated four times to

create four csv files; containing 1500 training examples representing four house types.

Table 8 Dataset development using Latin hypercube sampling

	Column 1	Column 2	Column 3	Column 4	...	Column 21
Row 1	0.9502	0.7208	0.5973	0.7063		0.7703
Row 2	0.6284	0.7032	0.3816	0.9435		0.6087
Row 3	0.7733	0.3287	0.0370	0.4914		0.9890
			...			
Row 1500	0.2981	0.4544	0.9825	0.8835		0.3642

Latin Hypercube Sample * (max -min)+min

	Depth	Width	WWR Front	RSI Exterior Wall	...	RSI Roof
min	8.035	3.637	0.008	0.652		0.479
range(max-min)	9.230	3.397	0.256	9.348		13.521

Output dataset of feature randomized house examples

	Depth	Width	WWR Front	RSI Exterior Wall	...	RSI Roof
House 1	16.8057	6.0852	0.1614	7.2549		10.8936
House 2	13.8350	6.0254	0.1062	9.4716		8.7092
House 3	15.1725	4.7530	0.0178	5.2454		13.8515
			...			
House 1500	10.7860	5.1802	0.2602	8.9111		5.4031

Figure 9 shows the distribution of some of the base feature inputs that were generated using *latin hypercube* sampling, note the even distribution of values throughout the range.
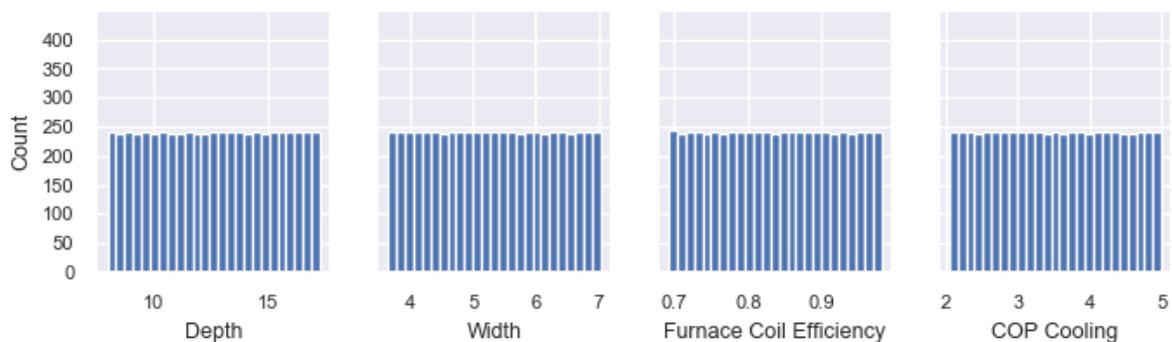
Figure 9 Distribution of features generated with latin hypercube sampling

The geometry of the baseline house models was modelled using an Openstudio [65] plugin for Sketchup [66] this geometry was exported into *Input data file (.idf)* form to be read directly in

EnergyPlus. The rest of the energy modelling was conducted with EnergyPlus directly,

Openstudio was only used to create the geometry of the baseline archetypes.

Using the *eppy.modeleditor* [67] package in Python, a script was developed to automatically

modify the baseline energy plus *.idf*s based on the input feature matrix values. Modifying most

EnergyPlus objects was very simple, but modifying geometry objects was quite complicated.

Creating a set of 6000 *.idf* files was one of the most challenging and time-consuming portions of

the research. Four different geometries with more than 100 coordinates and complicated

shapes were modified. A small sample of the code used to modify geometry can be found in the

appendix.

A range of sizes from small to large and a range of aspect ratios were used for each archetype.

Multiple geometries were modelled to make the archetype model more reflective of the

Toronto housing stock. The existing baseline archetypes were confined within the scope of the

neighbourhood that was targeted during the field study. The existing baseline archetype

approach uses only one house size per archetype. There is a range of house sizes within an

archetype, and one size will not necessarily describe every house if optimizations are

performed. A range of house sizes for each different archetype helps to address the small

sample size used to determine the baseline model for each archetype, and expand the

application of the current bottom-up archetype approach in Toronto. The same range of length,

widths and heights were used for each archetype. Although the "small" and "large" two-storey

semi archetypes had slightly different average baseline geometries, the difference was small,

and the sample size was also small. The small and large labels are representative only of the

baseline energy model, not the surrogate model. The same size range was used to enable

horizontal comparison between archetypes.

Modifying 6000 energy plus *.idfs* by hand would have been unfeasible, therefore writing a script

to do this automatically was required. The script had to apply to four distinct house shapes so a

script that could work on almost any house type shape was developed. In order to simplify the

process of modifying house geometry, all proportions on the x, y and z axes were kept constant

and multiplied by a factor based on the depth, width and height respectively. The house is

essentially stretched in the x, y and z axes. This created a range of aspect ratios.

	House a)	House b)
Width (m)	6.19	4.44
Ceiling Height (m)	3.08	2.69
Depth (m)	17.25	8.48
Above Grade Foundation Height (m)	0.81	1.11
Below Grade Foundation Height (m)	1.01	1.39

Figure 10 shows an example of two different small century semi samples used for training the

surrogate model.
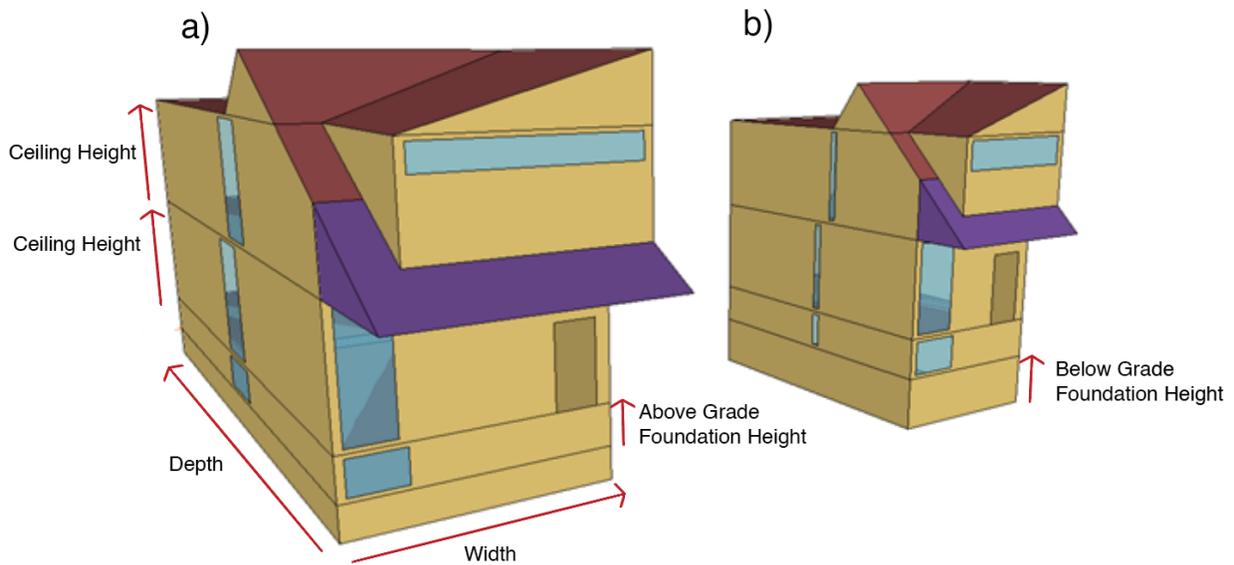
	House a)	House b)
Width (m)	6.19	4.44
Ceiling Height (m)	3.08	2.69
Depth (m)	17.25	8.48
Above Grade Foundation Height (m)	0.81	1.11
Below Grade Foundation Height (m)	1.01	1.39

Figure 10 Geometry modification example showing two differently sized house models used as training examples

The above examples show the geometry features that are modified for each training example.

House a) is an example of a house with a larger footprint, house b) is an example of a house

with a smaller geometry footprint. The width, ceiling height and depth parameters change the

proportions of the house. The base geometric proportions on each axis are maintained. An

example is that the length of the attic space in relation to the depth stays the same in both

houses while the depth changes. The height of the attic space in relation to ceiling height stays

constant between houses but the ceiling height changes. Width, ceiling height, depth, and

above and below grade foundation heights vary independently in each house example. The window to wall ratio also changes independently between houses.

Once the script was created, 6000 *.idf*s were generated based on the randomized feature matrix. The process of generating the full set of *.idf* files took less than a minute. The *.idf* files were split into four batches and run simultaneously by assigning each batch to one of the available CPUs. A script was written to extract the annual energy data from each output file. Each *.idf* file generated represents a single sample, the inputs used to generate the *.idf* are the features, and the target variable is the energy use.

3.1.2   Preprocessing

Based on best practices for model development recommended in [38], the dataset was split into training, validation and test sets for model development. The training set was used to train the model; the validation set was used to tune hyperparameters and select features, and the test set was used to test the model and report the accuracy.

Train/Validation/Test Splits

Nested Kfold cross-validation was implemented with *sklearn.model_selection.KFold* [56] Nested cross-validation was used because one of the objectives of this research was to see if a generalized combined model can achieve the same accuracy as individual models. Mean prediction accuracy offers a better comparison point than a single estimate of accuracy. Nested cross-validation refers to the use of an inner cross-validation loop within an outer cross-validation loop. Nested provides a mean test accuracy for a model and offers insights into the

model. As an example, the stability of the selected model can be determined by comparing the models selected between different splits.

The entire dataset was split into five equal subsets using 5-fold cross-validation. One subset was used as the test set, and the remaining subsets were used for training/validation. Each of these subsets is referred to as an outer fold. The outer 5-fold cross-validation split determines the test set and training/validation sets. In each outer fold, the training/validation set refers to the remaining four folds of the data that do not include the test set.

For each outer cross-validation fold, an inner cross-validation was performed with the training/validation by splitting this set into ten equal subsets with 10-fold cross-validation. Each of these subsets was used exactly once as the validation set while the other subsets were used for training. Once the optimal set of hyperparameters was chosen, the model was retrained using all the data in the training/validation set and evaluated on the test set. Including the data from the validation set in the training set increases the number of training examples the model sees and will improve the accuracy of the model when tested on the independent test set. This inner cross-validation is performed for each fold of the outer cross-validation loop.

The cross-validation methodology used in this research is explained further in Figure 11.
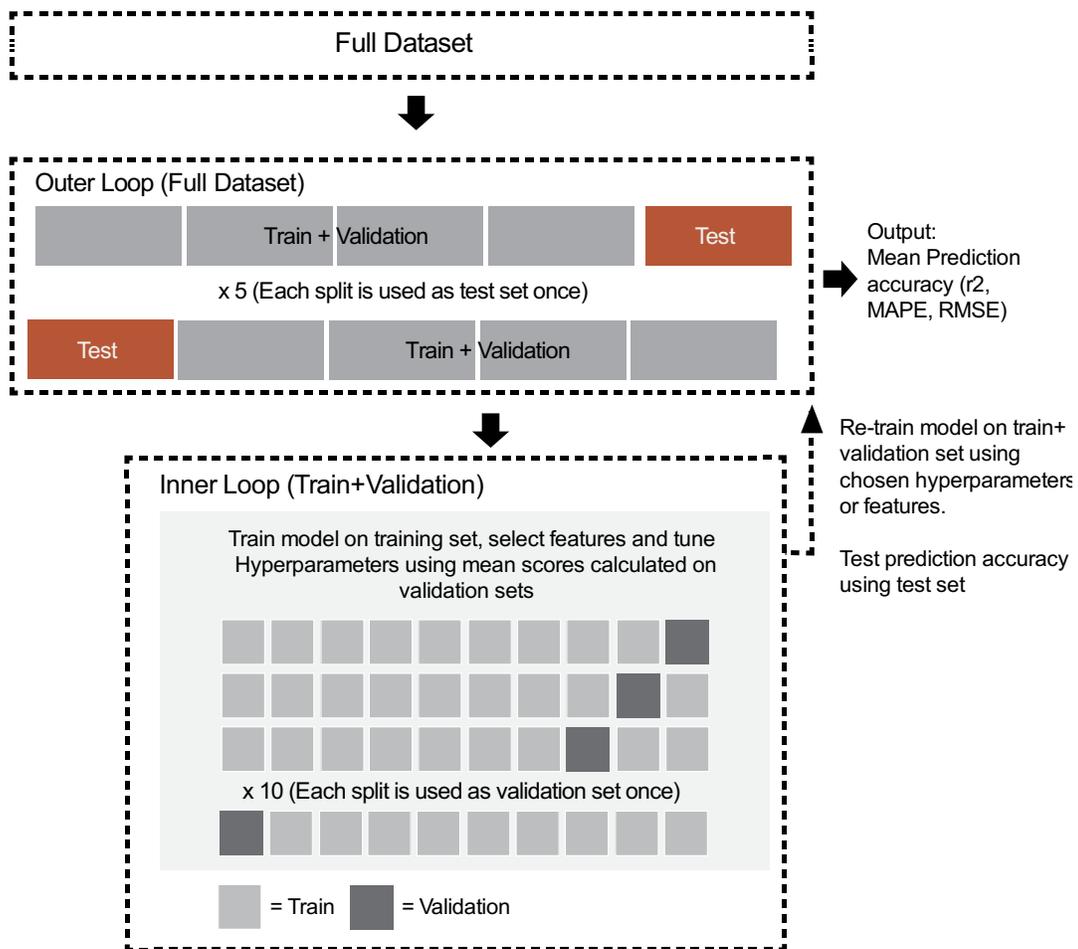
Figure 11 Nested cross-validation methodology.

The result of nested cross-validation is five different models, as different features and

hyperparameters can be chosen within the inner loops for each outer loop. Multiple splits allow

for the mean test scores to be estimated. Test scores generated from one split might be an

optimistic estimate of true test scores, so training on multiple splits was required. The model

was fully retrained, and stepwise feature selection was performed for each outer loop.

When the same features or hyperparameters were selected in each outer fold, the model was considered stable. If the selected model varied substantially between outer cross-validation loops, this was a sign that more samples were required. Unstable model can be difficult to interpret because the features or hyperparameters selected differ based on the split of data selected for training.

Feature engineering

Four different transformations ("non-transformed", "reciprocal", "logarithmic" and "square root") were applied to the input and output features to determine what transformation would result in the best fit to the validation set Testing all potential combinations of these transformations to the input and output features resulted in 16 unique combinations of input/output feature transformations. The mean cross-validation R^2 scores were used to compare the accuracy and select the best model.

Input features were standardized so that the feature values had a mean of zero and a standard deviation of one, using *sklearn.preprocessing.StandardScaler*() [56] . Standardizing the data is an important step that puts the features on the same scale and allows for sensitivity analysis using the feature coefficients. Standardizing is also required if regularization is used [38].

The mean and standard deviation both contain information about the test set. When the dataset is split, the test set can no longer be considered a set of unseen samples because the mean and standard deviation have been used to standardize the dataset. The standardization step should occur within each split of the cross-validation to avoid data leakage.

A script was written with the help of *sklearn.Pipeline* [56] to simultaneously perform the required preprocessing steps on the categorical and numerical features and then fit the data to the desired estimator, with less risk for accidental data leakage.

3.1.3 Creating Features for the combined model.

No research could be found on the development of features for use across multiple building types. Using a single model for multiple houses adds a level of complexity to determining features to investigate. The same set of features may not be defined in each house type.

Often the base inputs on their own do not yield high enough accuracy, combining features to create new inputs can result in an improvement to model performance. A set of 37 additional geometry features was created to find features that can be used to predict energy use more accurately than the base feature inputs on their own. For example, instead of window to wall ratio, window area was investigated in addition to window area per orientation. In total, including the base feature inputs 45 geometry feature options were considered. Additional features values were calculated during the creation of the dataset with *eppy*. These features are not just simple combinations of input features due to the geometry of the houses and the way the dataset was developed.

The goal when creating features was to describe the geometry in multiple ways individually and then in increasing levels of generalization. As an example, depth or height are more specific, above-ground wall area, and volume are more general geometry features.

The distribution of numeric feature values across house types is an issue that was not anticipated. Input features should either be continuous numeric values or categorical features.

Combining these two can create a complex problem. Examples of semi-continuous features include some Exposed Wall Area features. In semi houses, the right side of the house is a shared wall (party wall) that is not exposed to the outdoors, so the area will be 0. In detached houses, this wall will be defined as a continuous numeric value determined by the range. In order to include semi-continuous features in the combined model, the effect of combining the wall areas on both sides of the house was investigated as a way to produce a continuous input out of two semi-continuous inputs. Although these features could be excluded to eliminate any issues with these features, this would mean that geometry features would not capture the effect of the additional exposed wall in the case of the detached house. Because combining semi-continuous features created continuous features, these features were included in the combined model.

Some features, such as wall front area and wall back area, were almost completely collinear in some archetypes. Adding these highly collinear features together created continuous input features that did not have strong collinearity to each other. Window to wall ratios were set to be different for each side of the house, as there was no risk of collinearity within window features.

Orientation originally included eight categorical inputs in order to describe 0, 90, 180 and 270 degrees in addition to 343, 73, 163 and 253 degrees in order to capture the tilt of Toronto's grid 17 degrees west from true north. The 17-degree tilt did not significantly change the energy use compared to a north aligned grid, so 343 and 0 degrees were grouped into a feature called "North", 90 and 73 degrees were called "East". South and west were created in the same way. The reference orientation is shown below in Figure 12.
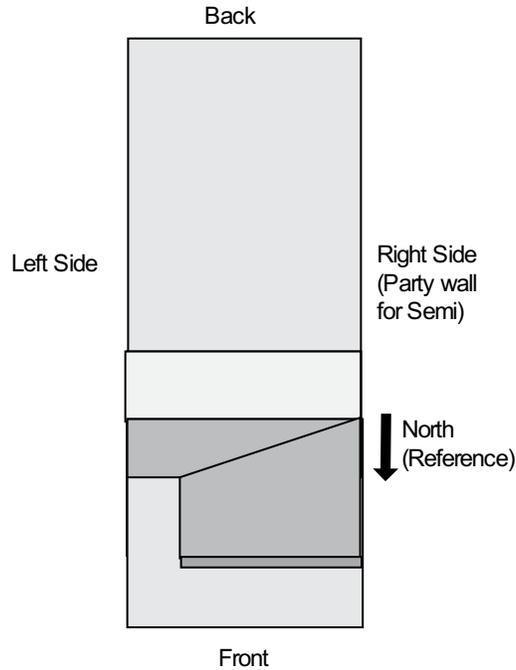
Figure 12 Reference building orientation

Several different subsets of features were included in the stepwise feature selection analysis

depending on the subset of archetypes being investigated. In Table 9 the features options that

are included for the feature selection analysis in all the archetype subsets (semi, detached and

combined) are shown. The + refers to two features being added together in a linear

combination in order to address collinearity or non-continuous behaviour.

Table 9 Feature options considered in stepwise regression, for all archetype subsets.

Semi, Detached and Combined			
Base Features	Lengths/Heights	Combined Wall Areas	Generalized Geometry Descriptors
U-factor Windows	Depth	Window Area North + South	Roof Area
Solar Heat Gain Coefficient	Width	Window Area East + West	Total Exposed Wall Area
RSI Slab	Above Grade Foundation Height	Wall Area North + South	Below Ground Foundation Area

RSI Foundation Wall	Below Grade Foundation Height	Wall Area East + West	Slab Area
RSI Exterior Wall	Building Height	Window Side Area Left + Right	Total Below Ground Area
RSI Roof	Ceiling Height Floors 1,2	Wall Side Area Left + Right	Conditioned Volume
Air Changes per Hour		Wall Front + Back Area	Surface to Volume Ratio
Furnace Coil Efficiency			Total Exposed Surface Area
COP Cooling			Above Ground Exposed Surface Area
Light Type (LED, Incandescent, Fluorescent)			Conditioned Floor Area
HRV Efficiency (0%,60%,85%,90%)			
Orientation (North, East, South, West)			
Neighbouring House Distance (1m, 10m)			

The above features are subdivided into columns based on their descriptions. The features that are only included in a specific subset of archetypes are grouped in Table 10.

Table 10 Feature options that are only included in a single subset of archetypes.

Semi and Detached Only	Detached Only	Combined Only
Window Side Area Left	Window Side Area Right	Number of Conditioned Floors (Two storey, Three-storey)
Window Front Area	Wall Side Area Right	House Type (Semi, Detached)
Window Back Area	WWR Right Side	
Wall Side Area Left	North Window Area	
Wall Back Area	East Window Area	
Wall Front Area	South Window Area	
WWR Front	West Window Area	
WWR Side Left	North Wall Area	
WWR Back	East Wall Area	
Wall Front + Back Area	South Wall Area	
	West Wall Area	

Base features are the original inputs to the model that are not related to geometry. These features were not included in the feature selection process because they do not have collinearity issues or geometry components.

Semi Feature Set

Semi-Detached Houses did not include any features that described the right side of the house on its own. The right side of the house is a party wall. This means the features describing this wall will not be defined, individual wall areas defined by orientation could not be included because when the party wall is facing a direction, the area at that orientation will be zero. Combined areas and combined orientation areas were included because these features did not include any 0 values.

Detached Houses Feature Set

All features were defined for detached houses because they have no shared walls. Categorical features describing the number of floors and type of house were not included because these options only apply to the combined dataset.

Combined Dataset Feature Set

The combined Dataset contains a mix of two and three-storey semi-detached and three-storey detached houses. Features describing specific walls of the house were not included because different walls of the houses are defined between detached and semi-detached houses. Only generalized geometry terms and combined terms were included in the model in order to answer the question of how a generalized model can be used to describe the housing stock.

In addition, the features chosen for the combined model are features that are general enough that they can be defined in all future archetypes considered.

The stepwise regression approach that will be used to select the features listed above is outlined in the following section.

Stepwise Multivariate regression

Some preliminary work has been conducted by Barnes [30] and Vu et al. [45] to develop a feature selection methodology using stepwise and lasso regression. An opportunity to add value to these existing methodologies is proposed through streamlining and simplifying the process, investigating collinearity in more depth, and addressing potential shortcomings. A stepwise selection process was developed. Elastic net was also used to create a model to compare the two feature selection methodologies and identify the pros and cons.

One of the issues in stepwise regression is that collinearity issues are not addressed or can be worsened [44]. Barnes addressed this issue by removing features that were known as linear combinations of other inputs in the model. Later in the process, univariate correlations were examined, and issues collinearity was detected still. This suggests that a more thorough analysis of collinearity should be addressed if stepwise is used.

A novel approach was developed to add a VIF threshold to a forward stepwise selection process. Often features are removed by filtering based on a VIF threshold prior to running a regression. In the case of a building energy model, combinations of geometry features will all be highly correlated because they can be calculated using linear combinations of other features. The challenge is determining which combination of features will improve the model performance without inflating collinearity to the point that the model can no longer be interpreted.

An opportunity to improve the methodology proposed by Vu et al. [45] was identified. This methodology looked at one rule at a time to select features, VIF value was targeted until a

threshold was reached, and then p values were used to test significance. By removing features by the highest VIF value, this process likely inadvertently removed features in order of the strength of their relationship to the target variable [45].

Using forward stepwise selection and testing if each feature was below a VIF threshold of 5 as it was added eliminated the issues identified in the methodology proposed by [45]. In this approach, features that would have been removed first due to collinearity are instead added to the model first because they showed the strongest correlation to the target variable.

During forward stepwise selection a base subset of original input features that were highly correlated to the target variable and not correlated to any other features by design were set as the base inputs to simplify the modelling process. Instead of starting with 0 features in the model, the feature selection process started with these base features already selected. During preliminary analysis, these features were always selected, so this simplification was justified. These features are identified in the first column in Table 9.

Forward Stepwise selection was performed by creating a script that iterates through a list of feature options. Each feature is added to the base features to create a new intermediate subset of input features. The VIF score for each feature in the new subset is calculated by regressing each feature on all other features. The amount of multicollinearity at each step is calculated. The VIF value is dependent on the features included in the model so it must be done for each subset, and not in a filtering step before. If any VIF score is above the threshold when a feature is tested, that feature is not selected, and the next feature is tested. If the VIF criteria is met, this feature is added to the base set of features, and it is trained and validated using 10-fold

cross-validation. If the feature decreases the cross-validation error, then the feature is marked

as the top option; each feature in the list is tested in the same way, and the feature that

provides the most substantial reduction to the cross-validation error is kept. The feature that is

kept is removed from the list of options and added to the base feature list to create a new set

of base features. The remaining options are rerun through the process until the validation

accuracy stops improving, or the VIF threshold is too high to include new features. The

proposed methodology is explained in Figure 13.
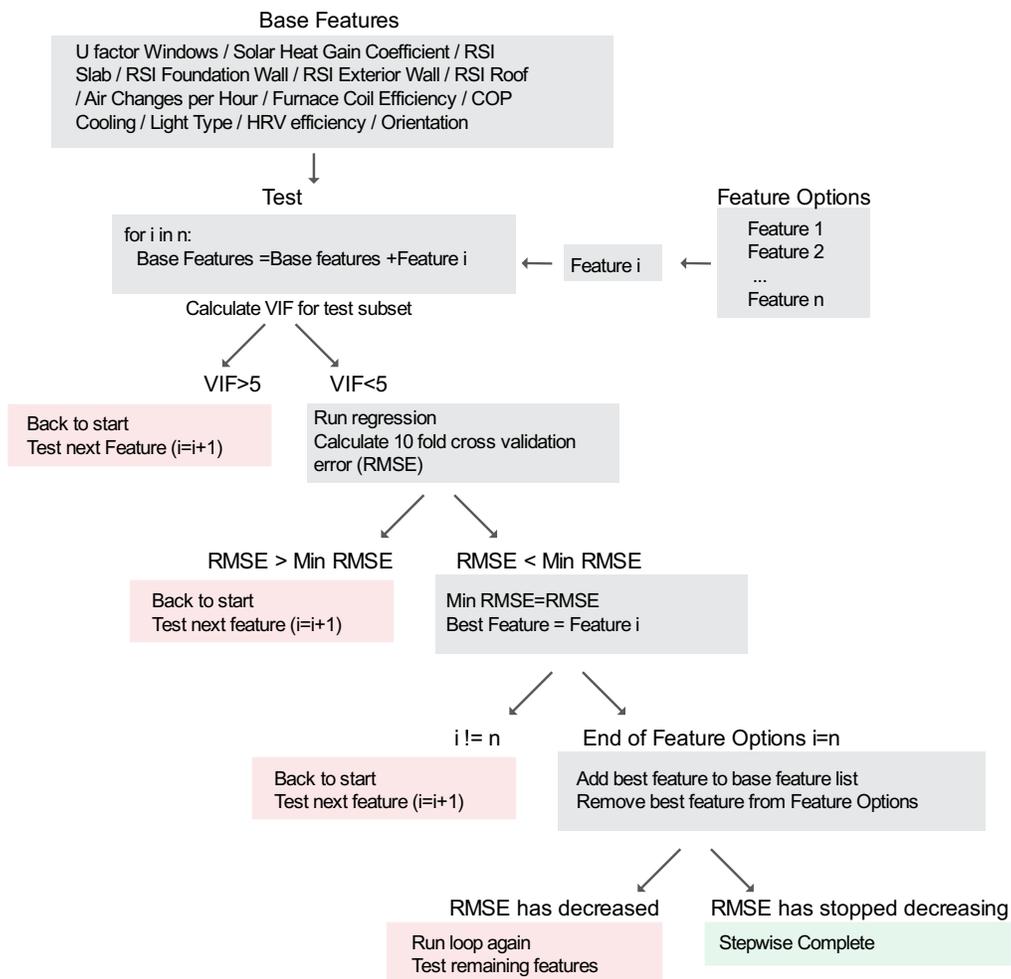
Figure 13 Proposed forward stepwise feature selection methodology

Nine different models were developed with stepwise regression in order to model and predict the target output for different subsets of the data. For each experiment, a model was trained using the stepwise methodology for each outer cross-validation split using the methodology outline. In total, 45 different models (9 models using 5-fold cross-validation) were created.

The first set of models was developed to predict energy use for each individual archetype, resulting in four models. The 1500 samples for each house were divided up into multiple testing training and validation sets as per the cross-validation methodology. Mean prediction scores for each house were recorded in order to be able to compare the accuracy between models. The Feature options used in stepwise selection for each house are listed in Table 9 and Table 10. The methodology used to develop the archetype specific models is outlined in Figure 14.

INDIVIDUAL ARCHETYPE MODELS

Develop and test four individual models using a subset of 1500 training examples representing each archetype

=10 Houses   Training   Validation   Test
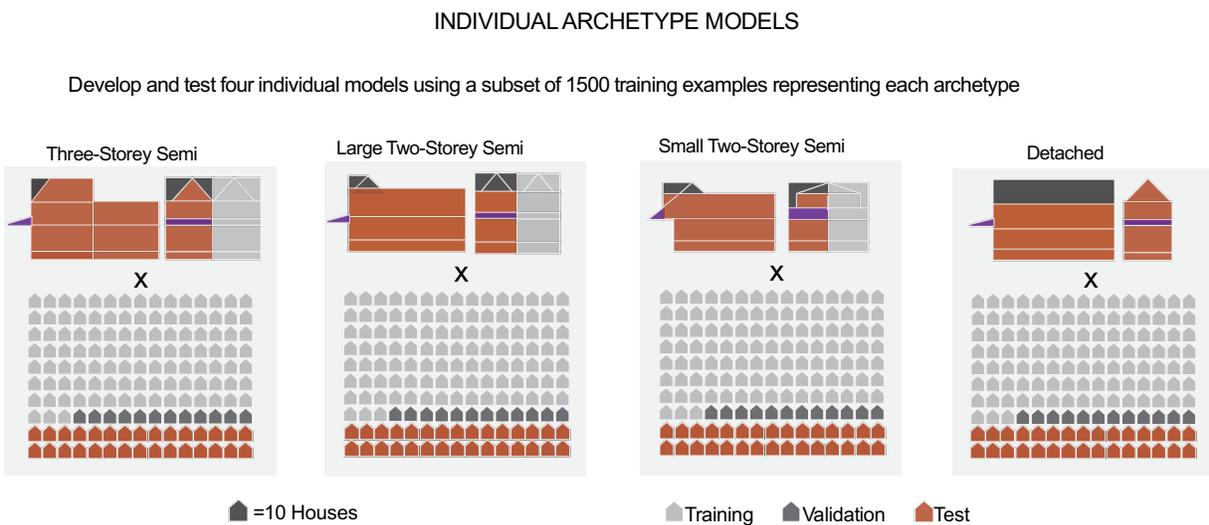
Figure 14 Training stepwise regression models on individual archetypes.

The datasets for each house type were merged to form a combined dataset representative of four archetypes. This dataset of 6000 houses was used to train and evaluate a new model. The feature options used for the combined model included only the generalized geometry

descriptors than for individual house archetypes because many features did not have continuous values when looking at multiple archetypes together. The set of starting features is listed in Table 9 and Table 10. The test set of 1200 samples was split into four parts in order to create a test subset composed of each archetype. Accuracy was calculated for each subset and on the entire subset. The combined model methodology is outlined in Figure 15.
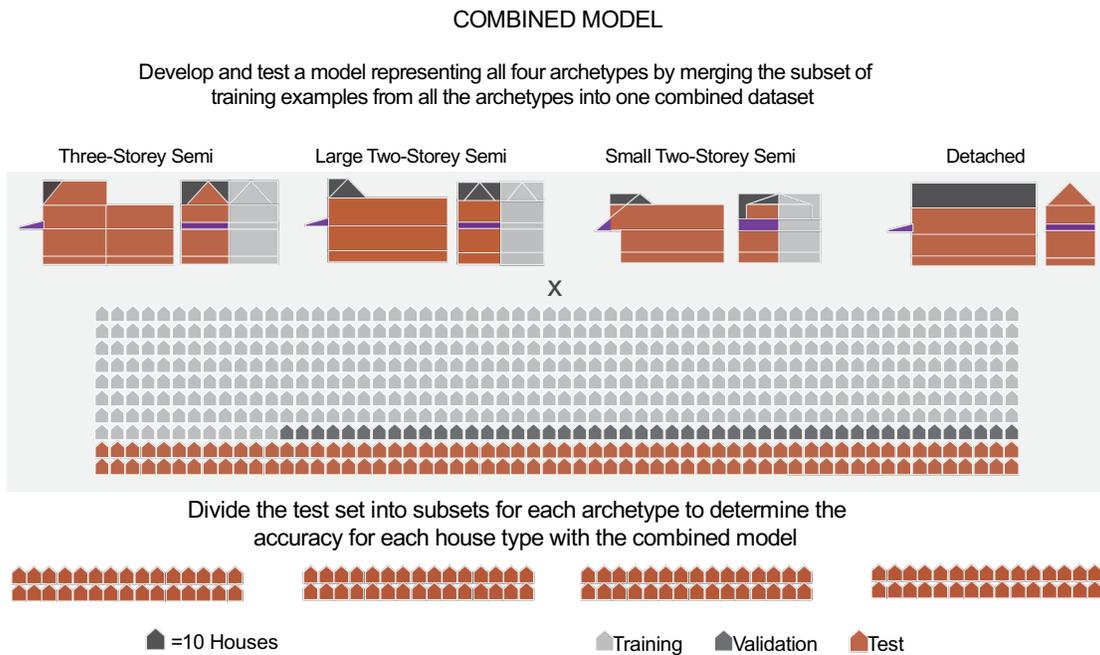
Figure 15 Training a model on combined dataset including all four archetypes.

The third set of models were developed to determine how well a model might predict on an unseen archetype. Creating a base energy model representative of every potential house geometry is not feasible, so the performance of a model, when tested on an archetype with a different geometry than it has been trained on, was investigated. For each model, three archetypes were used to train the model and perform the stepwise selection. The fourth archetype is used as a test set for the model that was developed. Nested cross-validation was

used to determine the mean prediction scores. The same set of feature options that were used

for the combined model was used in the stepwise selection for each group of houses. The effect

of the House Type category was also investigated to see if including house type (semi or

detached) increased the accuracy, if not, this would suggest that the difference between house

types can be captured entirely by generalized geometry features. The unseen archetype tests to

test the generalization ability of the model are described in Figure 16.
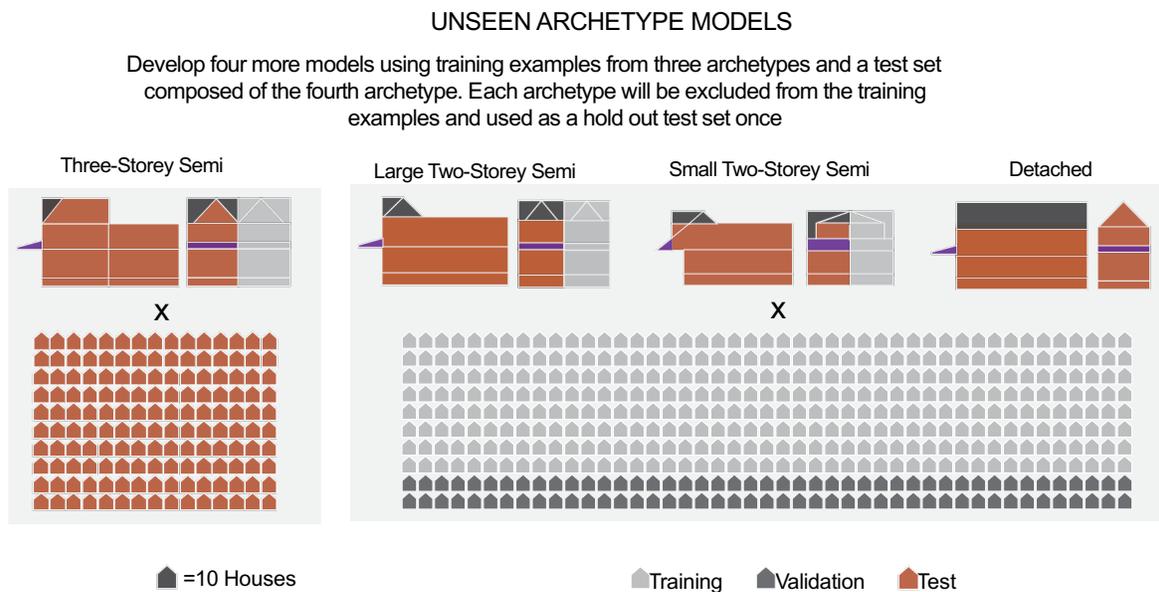
Figure 16 Unseen archetype tests to estimate generalization performance.

Figure 16 shows one out of four models that will be developed during the unseen archetype

tests. Each archetype will be used as the test set once for a model trained on the remaining

archetypes.

Embedded feature selection

Elastic net regression on its own as a feature selection tool on a large set of collinear features

has not been investigated in the literature. Elastic net helps reduce variance in the model by

introducing bias, which will help reduce generalization error. Elastic net is designed to reduce

collinearity through shrinkage of the coefficient estimates and elimination of some collinear

features [68]. Elastic net combines the behaviour of ridge and lasso regression. The L1

parameter determines how elastic net behaves. The L1 value falls between 0 and 1. 0

corresponds to ridge regression, and 1 corresponds to lasso regression. The tendency of lasso

regression to drop collinear features randomly was not desired, so a range of L1 values were

tested to determine the optimal characteristics. The choice of L1 parameter depends on the

application. It is more important to determine the optimal tuning parameter value. A sparser

model was desired; therefore, L1 values of 0.5, 0.65, 0.8, 0.95 were tested. The model was run

with a range of l from 0.0001 to .1 for each l1 value. Plotting the prediction scores compared to

the number of features is used to determine the desired tuning optimal tuning parameter value

by finding the point at which prediction accuracy starts rapidly decreasing. This graph shows

what potential accuracy is possible based on the number of coefficients to enable comparison

to the stepwise model. The elastic net model was compared to the stepwise model based on

the stability of the predicted coefficients and the number of features that are included in both

models at the same accuracy.

To compare the stability of coefficients between the model's, the standard deviation of

coefficient estimates were compared between the stepwise and elastic net regressions. The

elastic net regression shrinks the absolute value of coefficients so the standard deviation will be

different between the two models. In order to allow a comparison of coefficient stability on a

relative scale, the coefficients were plotted with standard deviation as an error bar for each

coefficient estimate. This allows a comparison of the magnitude of the coefficients relative to

the amount of deviation and helps diagnose the possibility for changes in the relative importance of features based on the range of potential estimates.

Artificial Neural Networks

The primary focus of this research is to develop a simple and interpretable model to predict energy usage for single-family houses. If interpretability is not a concern, it is helpful to know if a black-box approach could fit the data better and whether future research should take this approach. Current research suggests that accuracy can be higher when using an artificial neural network (ANN) model, especially if non-linearity exists in the data [12]. A preliminary analysis was carried out to determine what level of accuracy can be achieved using ANN. TensorFlow [69]and Keras [70] were used to develop an artificial neural network, perform the analysis and build the architecture of the network. Tuning an ANN model is crucial to develop an acceptable accuracy and prevent overfitting [12]. Several important hyperparameters define the structure of an ANN and how it learns. The number of epochs which defines how many times the weights are estimated through backpropagation was selecting by adding a callback which stopped the training process when the validation accuracy stopped improving. The average number of epochs in which the model converged was used as the input for a model trained on the training and validation set and evaluate on the test set. Batch size, which defines how many samples are used before weights are updated, was set to 80 based on trial and error. Most researchers have only used one hidden layer and were able to achieve high accuracy without serious overfitting problems that can arise when multiple hidden layers are added. Seven different hidden node configurations were tested using cross-validation to determine where the validation scores reached a minimum, and the validation error started increasing. The residuals

and predicted vs. actual energy use plot and performance metrics were used to compare the fit

of the ANN to the fit of the models generated with regression techniques.

4 Results and discussion

Phase 1 field research and energy modelling baselines

The results of phase 1 are summarized in this section. Phase 1 included:

- the results of a field study conducted in an urban Toronto neighbourhood

- the creation of new baseline energy models and updates to existing models

- Development of ranges

- Dataset development

4.1.1   Data Collection and Determining Archetypes

Collecting data from houses and determining baselines was conducted in 2019. Eighteen

houses were measured, 13 of these were used directly to determine ranges for baselines.

Some houses were measured and then not included based on the findings. For example, a few

houses were found to have major renovations, which made taking average measurements

difficult. Initially, an additional three-storey semi was included, the measurements showed the

dimensions were similar to the existing three-storey model. Multiple two-storey semis were

initially included in different sub-groups.

Data collection results for the two-storey semi archetypes are included in the appendix.

Small Two-Storey Semi

Figure 17 Figure 18 show two subsets of small Two-storey semis that were grouped separately

during preliminary surveying work.

Figure 17 Small two-storey semi houses in the Pocket [52]

Figure 18 Small two-storey semi houses without peaked roof detail [52]

Although they appear different from the front, the dimensions are very similar, and the conditioned spaces have the same geometry. The only difference was found in the amount of unconditioned attic volume; the comparison between roof shapes of small two-storey semis can be seen in Figure 20. This small increase in unconditioned attic space would not have a significant effect on energy demand, so these house types were grouped into one archetype. This archetype is prevalent throughout urban Toronto neighbourhoods based on a preliminary observational study. Some neighbourhoods in Toronto, including Chatham Avenue, In the Pocket neighbourhood, are predominantly made up of this house shape.

The small two-storey semi was defined as the small semi for clarity purposes. This house was only slightly smaller than the large-two storey semi based on the field study results. This house is characterized by a cantilevered second floor with a large sloped roof extending from the attic to create a porch overhang. The floor to ceiling height was constant across the second floor in all houses measured. The area under the peak roof and above the ceiling was unconditioned attic space. In some cases, there was a division between the attics of both houses. In other cases it was a continuous attic space between the shared houses. The second-floor overhang and jog in creates a greater exposed surface area in this house compared to a simple rectangular shape. The shape of the roof is very different than other houses measured, this is a very common design throughout Toronto.[52]
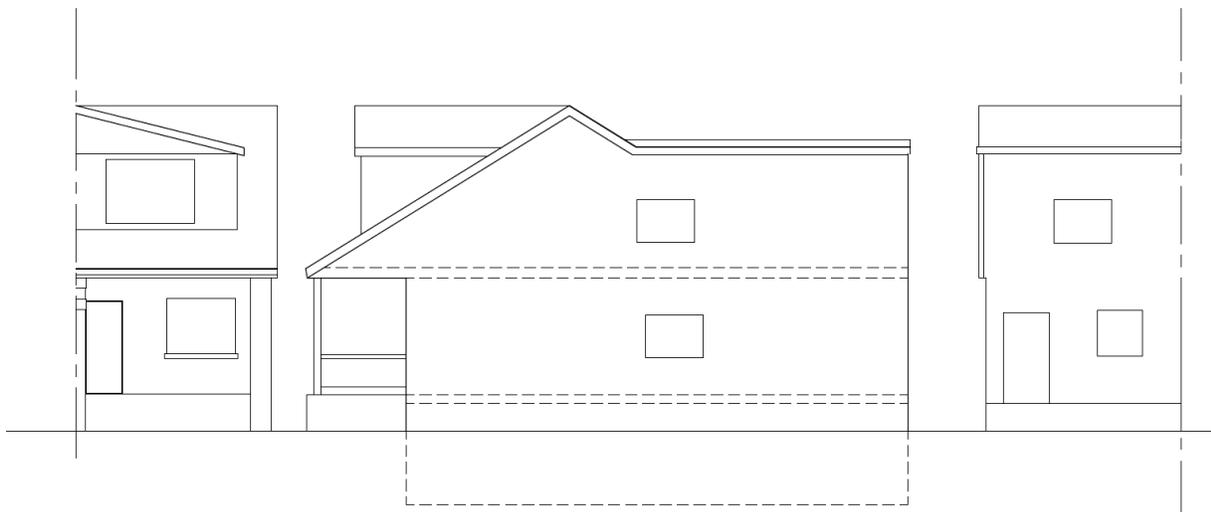
Figure 19 Small Two-Storey Semi peak roof.
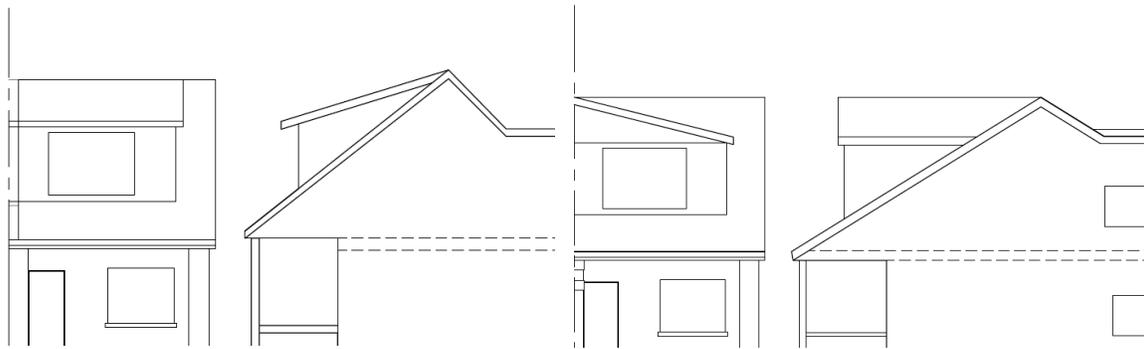
Figure 20 Small two-storey semi, comparison of roof details.

The baseline models generated using the OpenStudio plugin for Sketchup are shown in Figure 21.
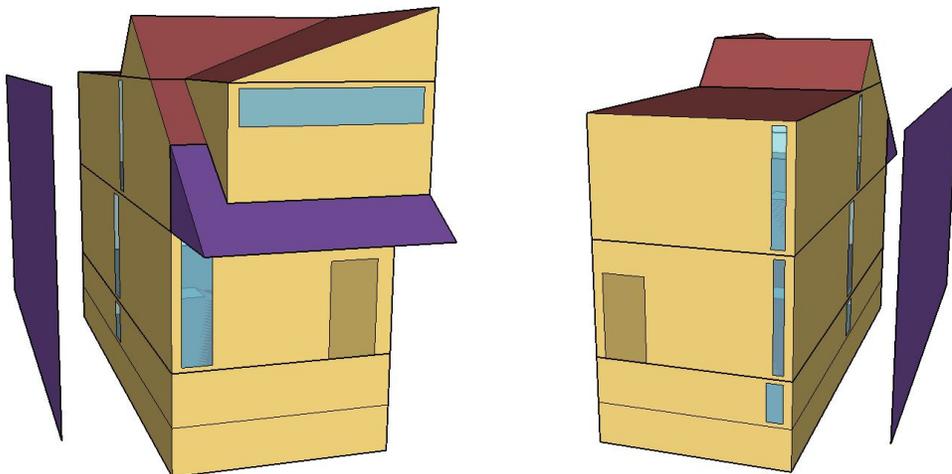
Figure 21 Small Two-Storey Semi OpenStudio model. (front & back)

the purple surface plane on the side of the house is a shading surface that captures the effect of a neighbouring house. A shading overhang surface is used over the porches of all houses. The third floor zone for the two-storey semi houses is an unconditioned attic space.

An example floor plan is shown in Figure 22. In many cases, the main floor had been converted into an open floor plan while the upstairs was divided up into bedrooms.
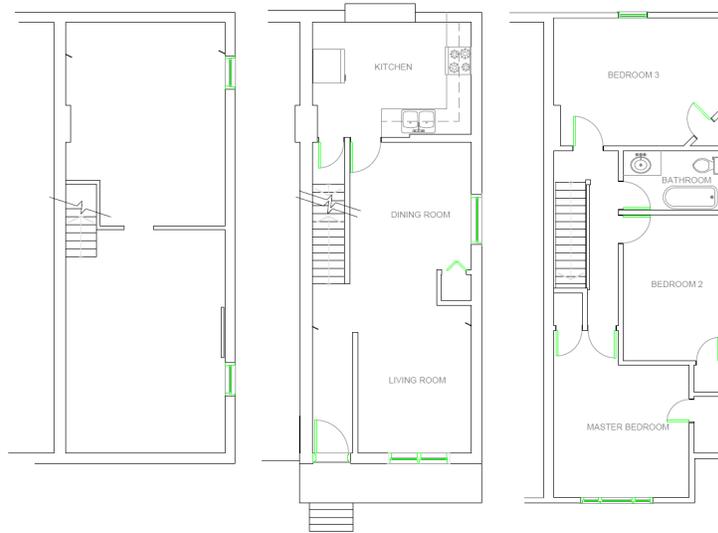
Figure 22 Example floor layout in a small two-storey semi.

Large two-storey semi

The large two-storey semi archetype was common throughout the Pocket neighbourhood; these houses were easily grouped and identified. Some large two-storey houses had attics with a peak above the front window, while others did not have this feature. The area beneath the roof was an unconditioned attic space in all houses. These houses have a simple rectangular layout and geometry.
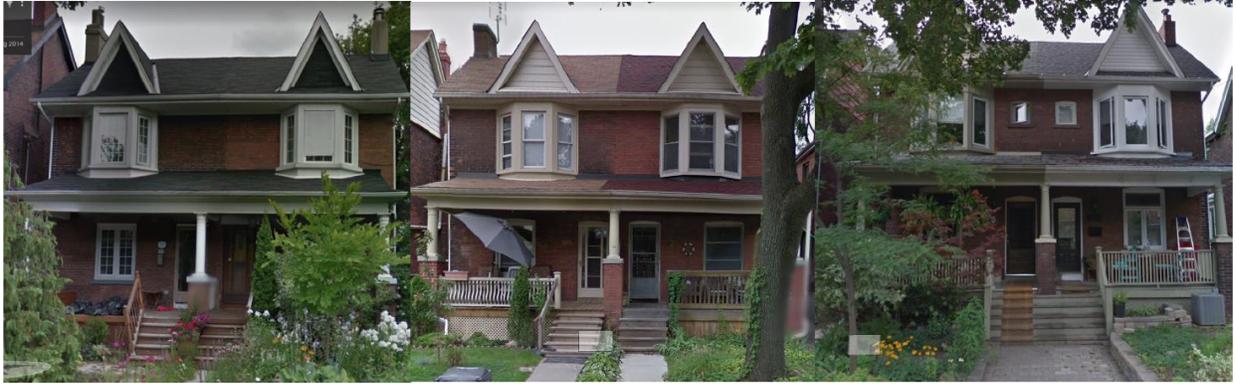
Figure 23 Large two-storey semis found in the Pocket [52]
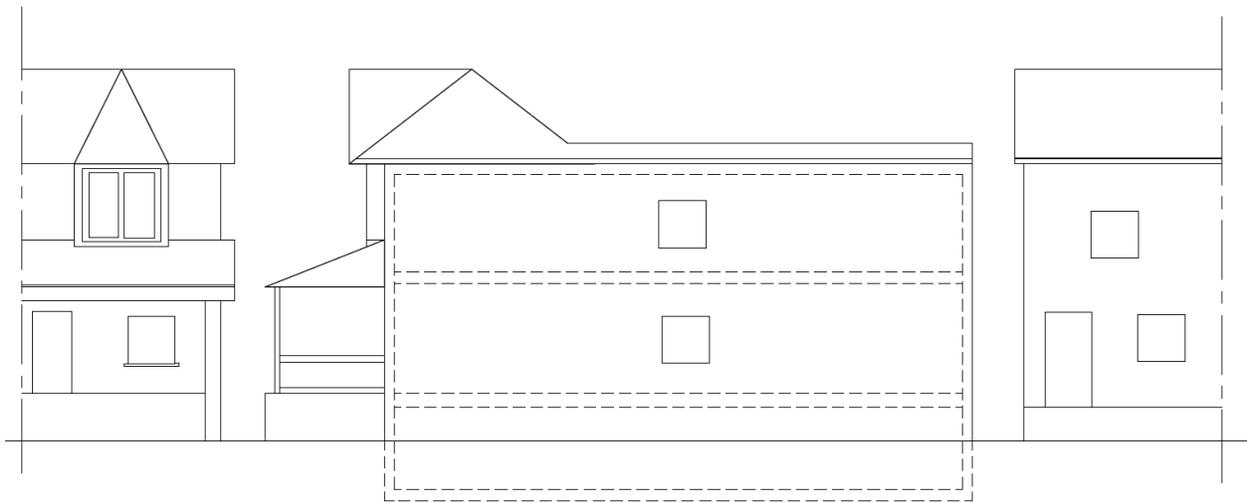
Figure 24 Large two-storey semi.
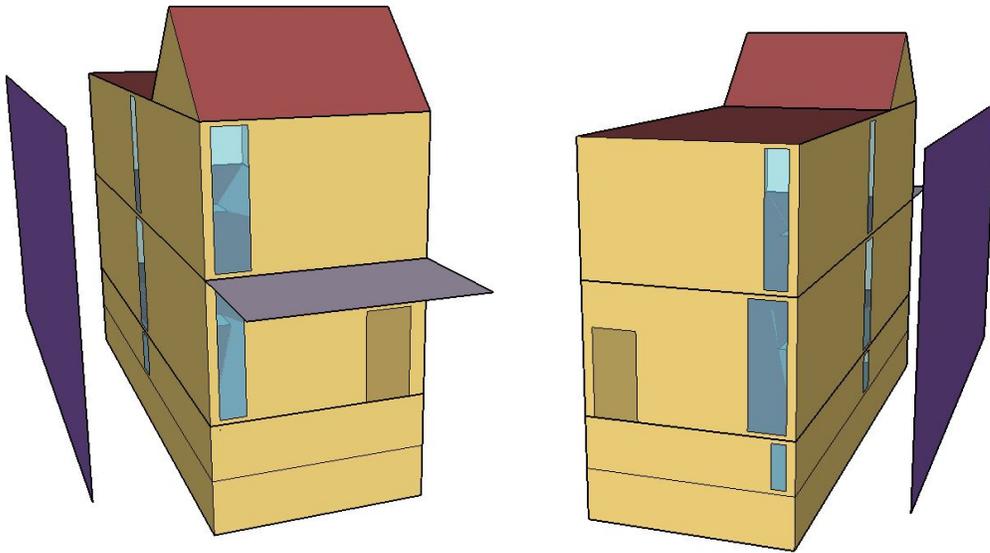
Figure 25 Large two-storey semi OpenStudio model (front & back)

Three-storey Semi

The three-storey semi is defined by having a conditioned third floor, sometimes the roof space

was extended to increase the area underneath the roof on the third floor to increase living

space. The third floor zone does not extend the length of the house.

Figure 26 Three-storey semi in Riverdale, Toronto [52]

The open studio baseline model developed by Jermyn[3] is shown below.
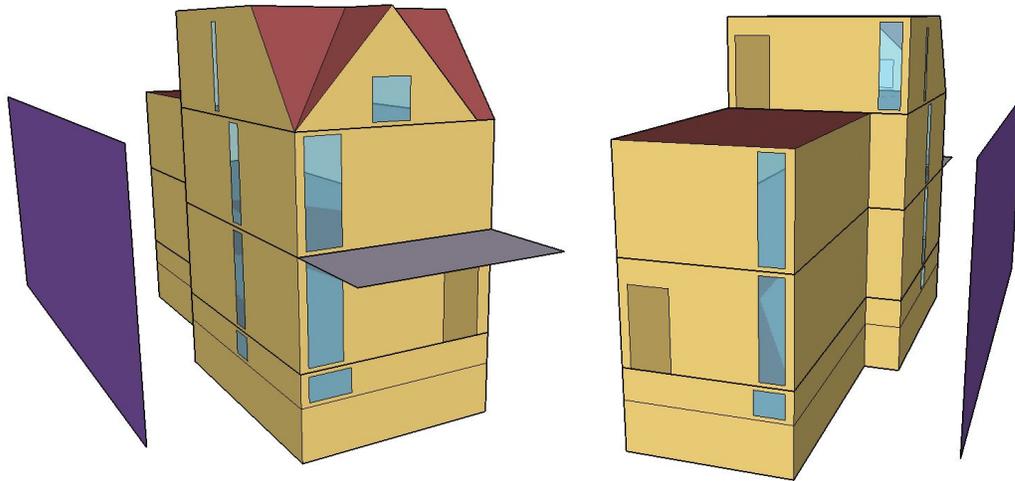
Figure 27 Three-storey Semi baseline developed by Jermyn [3] *(front & back).*

Detached

The detached archetype has 3 floors and a rectangular geometry. This archetype does not share

any walls with neighbours, the third floor zone is located in the area underneath the roof.

Figure 28 Detached baseline developed by Jermyn [3].

A summary of the baseline dimensions for each archetype can be found in Table 11 Baseline

dimensions for each house type

Table 11 Baseline dimensions for each house type.

	New Research Contribution		Existing Research	
	Large two-storey Semi	Small two-Storey Semi	Three-storey Semi	Three-storey Detached
Plan Shape	Rectangle	Rectangle	L-Shape	Rectangle
Depth	13.3m	11.6m	13.7m	13.4m
Width	4.7m	4.57m	5.1m	5.8m
Ceiling Height Floors 1,2	2.7m	2.6m	2.75m	2.75m
Basement Height Below Grade	1.2m	1m	1.3m	1.4m
Basement Above Grade Height	1.2m	1.1m	0.8m	0.9m
Heated Floor Area	187.5 m2	164.0 m2	225.21 m2	250.9m2
Glazing % Front	12.6	13.4	15.6	19
Glazing % Left	4.9	4	3.6	5.7
Glazing % Right	-	-	-	3.2
Glazing % Back	11.9	5.7	10.8	15.6
Reference:	Field Study		Jermyn [3]	

In addition to the measurements, a survey was conducted. The survey can be seen in the appendix. The results of the survey suggested that cost was the largest barrier to carrying out retrofits in their homes, this reinforces the need to find the lowest cost retrofit solutions possible. A scale of 1 to 5 was used with 1 representing no barrier, and 5 representing a substantial barrier to performing a deep energy retrofit. On average cost received a 3.15, aesthetics received a 2.94. Unclear savings goals also received an average of 2.76. Unclear saving goals ranks as a substantial barrier to performing retrofits. This finding also reinforces the need to quantify the potential savings for each house through modelling.

Residents were also asked how much energy reduction they would hope to achieve with retrofits and what their budget would be, the results are included in the appendix. On average, residents hoped to reduce energy use by 46.43% for a cost of $ 33600. This represents unrealistic energy savings goals given the budget expectations. Clearly quantifying the steps

required to achieve an energy performance target and the cost is important for homeowners considering retrofitting their homes.

4.1.2 Updating Energy Plus Model

The existing baseline archetype energy models created by Jermyn provided a model from which to build the models that were used in the surrogate model. Jermyn's[3] models were calibrated to the monthly energy usage of a typical house by adjusting occupant behaviour.

Although these models were able to be used to accurately predict the energy usage of a fixed baseline archetype house, they were not suitable for use in a surrogate modelling application.

The IECC prototype single-family models were used as a reference for updating the baseline models. The most significant changes to the model included: auto-sizing HVAC components, simplifying the window model, updating the mechanical ventilation/heat recovery system and changing the ground heat transfer model.

System Sizing:

In Jermyn, fixed values were used for HVAC sizes across Century Detached, Century Semi and in all retrofit cases analyzed. This approach will create a system that is likely to be oversized or undersized if parameters affecting building performance are modified. Certain zones may be significantly above or below setpoints temperatures, especially in retrofit cases. These single-family homes only have one control zone, which makes it difficult to meet setpoint temperatures in all zones. Proper sizing of heating coils and airflow rates will result in a building that is more accurately modelled across all zones. In the updated baseline models, the heating and cooling system was fully auto-sized following the approach used in the IECC prototype

buildings [72]. *Ann Htg 99.6% Condns DB* and *Ann Clg .4% Condns* were used as the design day

objects for sizing calculations. The system heating and cooling airflow rate and coil capacities

were sized on the total sensible load of all zones. The fraction of the total heating airflow going

to each zone Air terminal was sized based on the sensible load in each zone.

Windows

The window construction object created by Jermyn was replaced with

WindowMaterial:SimpleGlazingSystem. This window object approximates a complex fully

specified window and frame using only Solar Heat Gain Coefficient and U-Factor. Switching to a

simple window allows for easy modification of window performance values with *eppy*. The

simple window construction has been shown to effectively model a more complex window in

EnergyPlus and is used in the IECC models for this reason [72]

Basement Model

Jermyn's [3] model used a slightly outdated approach to basement heat transfer calculations

that requires the use of an auxiliary software tool called *Basement*. *Foundation Kiva*, is a more

recent alternative to *Basement* which approximates these calculations within 3% mean

absolute standard deviation from the values calculated with *Basement* and takes only seconds

[73]. This is a simplification of the modelling process that will simplify the process of generating

.idf files and maintain a high level of accuracy.

Mechanical Ventilation

The IECC model does not include mechanical ventilation, Jermyn specifies a heat exchanger

with a constant flow of 0.05 l/s throughout baseline and all retrofitted cases. The

AirLoopHVAC:Unitary:Furnace:HeatCool object used to model the furnace accepts only on/off

or constant volume fan selections. An on and off fan is used in the IECC prototype and correctly

models the behaviour found in archetype houses. When fan onoff is used, the furnace fan and

heating/cooling coils are cycled on and off in unison in order to meet the setpoint in the control

zone. When the furnace fan is on mechanical ventilation will be supplied at the set rate, when

the furnace fan is off no mechanical ventilation will be provided. The actual mean daily mass

flow rate of outdoor air is for the baseline model is shown in Figure 29. The amount of outdoor

air fluctuates throughout the year in relation to heating demand and total system airflow. The

ASHRAE 62.2 standard determines the required mechanical ventilation in a retrofitted house

based on building footprint, number of bedrooms and an airtightness credit. The amount of

mechanical ventilation required is inverse to the airtightness in order to ensure acceptable air

indoor air quality. In the most air-tight model presented by Jermyn, the mechanical ventilation

requirement should be the greatest, the results in Figure 29 show the opposite to be true. The

combined effect of an oversized heating system that is rarely turned on and a mechanical

ventilation system tied to furnace operation results in a negligible amount of outdoor air being

supplied to the building.
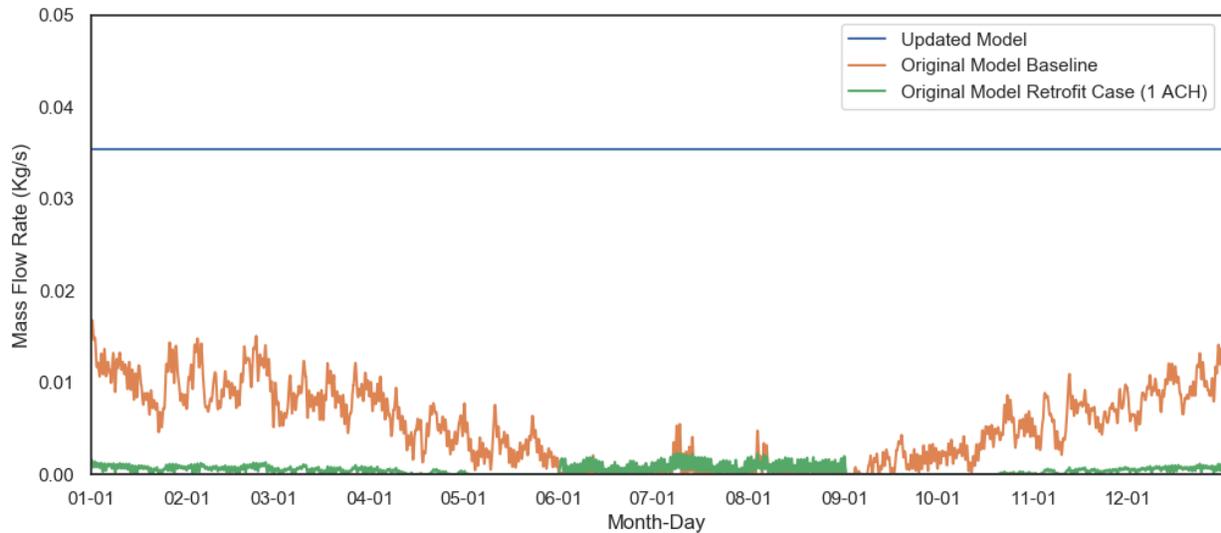
Figure 29 Mechanical ventilation supplied in various model iterations.

The existing heat exchanger object and outdoor air loop were removed from the model. In its place the *ZoneHVAC:EnergyRecoveryVentilator* was used as a zoneHvac object in the main floor control zone. This object delivers a constant outdoor airflow rate independent of the heating and cooling system operation. The airflow rate in the updated baseline detached model can be seen above in Figure 29, a constant mechanical ventilation rate is provided throughout the year.

The annual energy use and energy use intensity calculated using the original modelling approach and the updated approach were reported below for the two new two-storey century semi archetypes. The energy performance of the three-storey century semi and century detached is also reported. The original values for these existing baseline models differ slightly from the original values as reported by Jermyn, due to the use of an updated CWEC weather file.

The updated modelling approach yielded a consistently lower energy use and energy use intensity over the typical meteorological year. Floor area for EUI calculation is based on total floor area as calculated by energy plus (depth times width) not treated floor area. The simulated EnergyPlus results are shown below in Table 12 and Figure 30 , the shaded region in the table includes the results of the energy models developed and run by Jermyn [3].

Table 12 Predicted energy use for each archetype compared [3]

	Existing Model Methodology		Updated Model		% Difference Overall Energy
	Overall Energy(kwh)	EUI (kwh/m2)	Overall Energy(kwh)	EUI (kwh/m2)	
Small Two-Storey Semi	**34,728**	**192**	**34,416**	**190**	**0.90**
Large Two-storey Semi	**40,439**	**199**	**37,594**	**184**	**7.29**
Three Storey Semi [3]	43,419	198	**41,980**	**191**	**3.37**
Detached [3]	59,283	194	**53,769**	**176**	**9.75**

Figure 30 Energy use for baseline homes visualised

The total energy for the new Two-storey Semi archetypes is lower than the existing Three-storey archetypes. This result is expected based on the size difference and number of floors and

84

anticipated when compared to the existing archetype energy usage. The Small Two-storey Semi

has the lowest overall energy of the group at 33,536 kWh. It is interesting that the Two-storey

Semis have the highest EUI. The Detached archetype would be expected to have the highest EUI

because it has an additional wall, however it has the lowest EUI. The high EUI in the two-storey

semi case may be due to a larger unconditioned surface area per zone. Out of 3 conditioned

zones only one shares two surfaces with neighbouring zones (the main floor zone with the

basement and second floor), whereas the century Detached archetype has two zones that

share two surfaces with neighbouring zones (both the second floor and main floor). The

Detached Archetype also has a third floor with the equivalent floor area as the other zones but

a fraction of the volume due to the peak roof, this will result in significantly larger floor area

and not a significantly larger conditioned volume.

The percent difference between the original baseline and updated baseline ranges from 0.46%

for the Two-storey Semi 1 to 9.75% for the Detached house type. The large discrepancy

between the two detached models can likely be attributed to an improperly sized heating and

cooling system that is delivering too much heat to the third and second floors. The mean daily

temperatures across the four zones of the Detached Archetype in the original model can be

seen in Figure 31. The third-floor zone is consistently more than 5 degrees higher than the

setpoint. Each zone in this model is receiving the same airflow rate despite significant

differences to sensible load that would be expected between zones due to differences in

exposed surface area and volume.
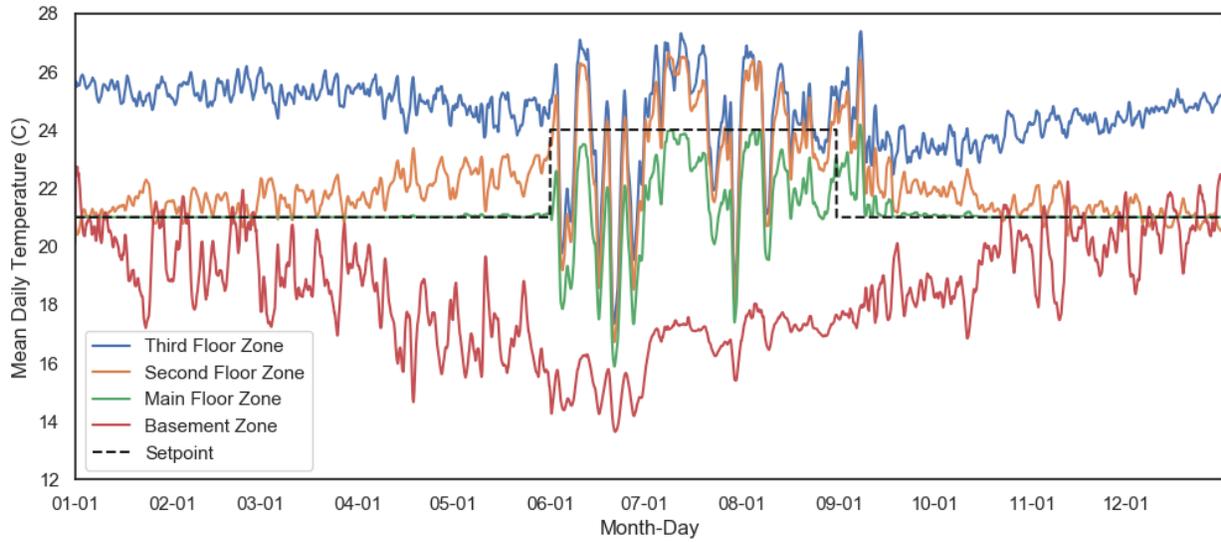
Figure 31 Zone temperatures in baseline detached model

The zone temperatures in the updated model are shown in Figure 32. Uncontrolled zone

temperatures are much closer to the expected setpoint temperature, and therefore, the

estimated energy usage for the updated model will be more accurate and lower than the

original baseline model. The same trends were noted for all updated models when compared to

the original model.
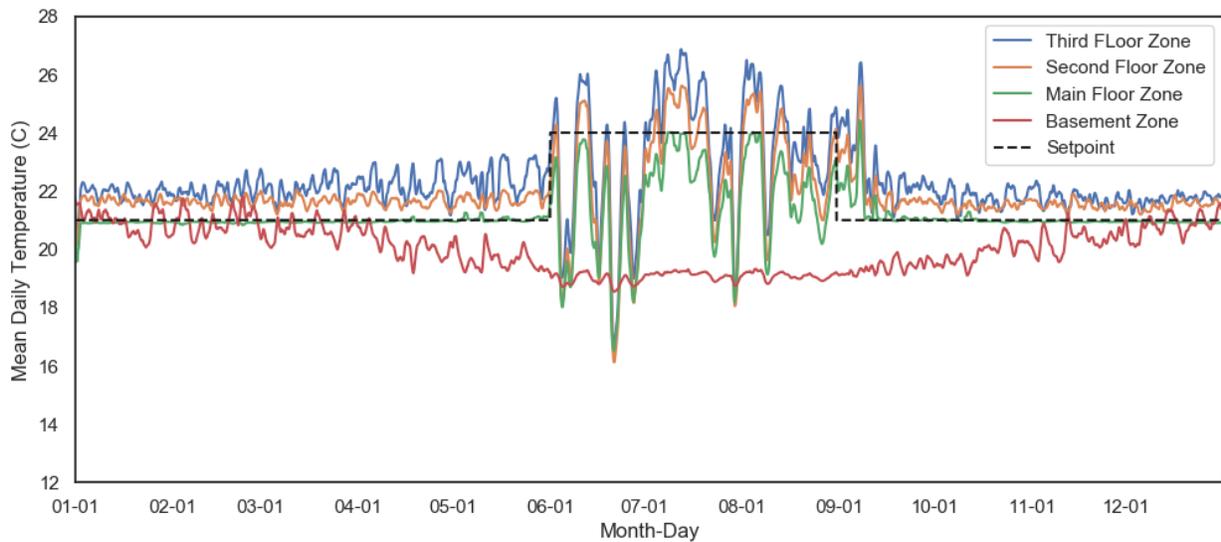
Figure 32 Zone temperatures in updated baseline detached model

86

The updated energy model using autosizing of HVAC components shows that the temperatures in the uncontrolled (slave zones) in the house are a much closer fit to the setpoint temperature, but they are still not meeting the setpoint temperature. This behaviour is inherent to the HVAC and control system used. Different house types, geometries and combinations of building features will result in slightly different temperatures in uncontrolled zones. Compared to a building where each zone is controlled, there will be a lot more inherent unpredictability in the annual energy use because the temperatures in each uncontrolled zones will vary slightly across houses. The temperature profile in the third floor zone is plotted for two houses with the same geometry but different insulation levels. The temperature is consistently higher in the insulated case on the third floor, even though the setpoint temperature is the same in both cases.
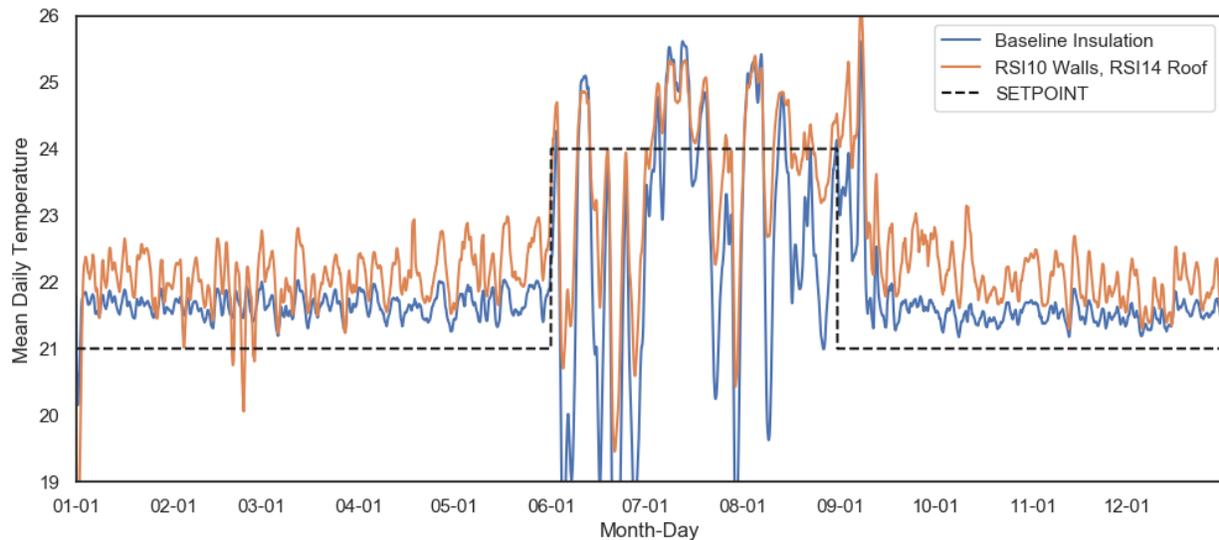
Figure 33 Third floor temperatures with varying levels of insulation.

The thorough analysis of the existing models, and the updates were essential in order to ensure the energy models were performing as expected. The surrogate model developed in phase 2 Is only as robust as the baseline energy model that is used to build the dataset. The results of the

field study were used to inform ranges of building characteristics that were then used to modify

the updated baseline energy models for each archetype. The result was a set of 6000 *.idf*s that

were used in phase 2 to build the surrogate models.

Phase 2: Surrogate modelling

The results of phase 2 are summarized in this section. In phase 2 the dataset of energy models

was used to explore the application of surrogate models for bottom-up archetype

development. This section begins by analyzing the results of the preprocessing and feature

engineering steps.

4.1.3   Transformations and preprocessing

Using only the base inputs and a linear model with no input or output transformations, the

model had poor accuracy (MAPE of 16.9% and R^2 of 0.76) and evidence of non-linearity. For

these reasons, several transformations were tested to see if a transformed model would

perform better. Three different transformations were tested to determine which would provide

the best fit to the data. These transformations were applied to the input features and target

features in every potential combination for a total of 16 different combinations of unique

input/ouput transformations. The 16 combinations included the untransformed inputs and

target. A sample of the transformations applied to the input and target features can be seen in

Table 13 The log input, and output transformation (referred to as a power transformation),

resulted in the highest R^2.

Table 13 Transformations of the input and target variables and their corresponding cross-validated scores

Input Transformation	Target Transformation	Validation R2	Validation MAPE

log10(x)	log10(y)	0.844	3.19
sqrt(x)	log10(y)	0.842	3.20
log10(x)	sqrt(y)	0.829	6.86
sqrt(x)	sqrt(y)	0.824	6.86
x	log10(y)	0.815	3.43
log10(x)	1/y	0.799	15.11
x	sqrt(y)	0.799	7.24
sqrt(x)	1/y	0.797	15.63
log10(x)	y	0.788	16.45
sqrt(x)	y	0.781	16.40
x	1/y	0.770	16.83
x	**y**	**0.757**	**16.95**
1/x	log10(y)	0.748	4.10
1/x	sqrt(y)	0.744	8.54
1/x	y	0.718	19.30
1/x	1/y	0.694	18.20

The untransformed input and output in addition to the best transformation are shaded in Table 13. The residual trends were also analyzed to determine the goodness of fit and verify that there are no trends in the residuals vs. predicted plots. The non transformed data can be seen in Figure 34. Note that "Actual Energy" refers to the energy use simulated with EnergyPlus.
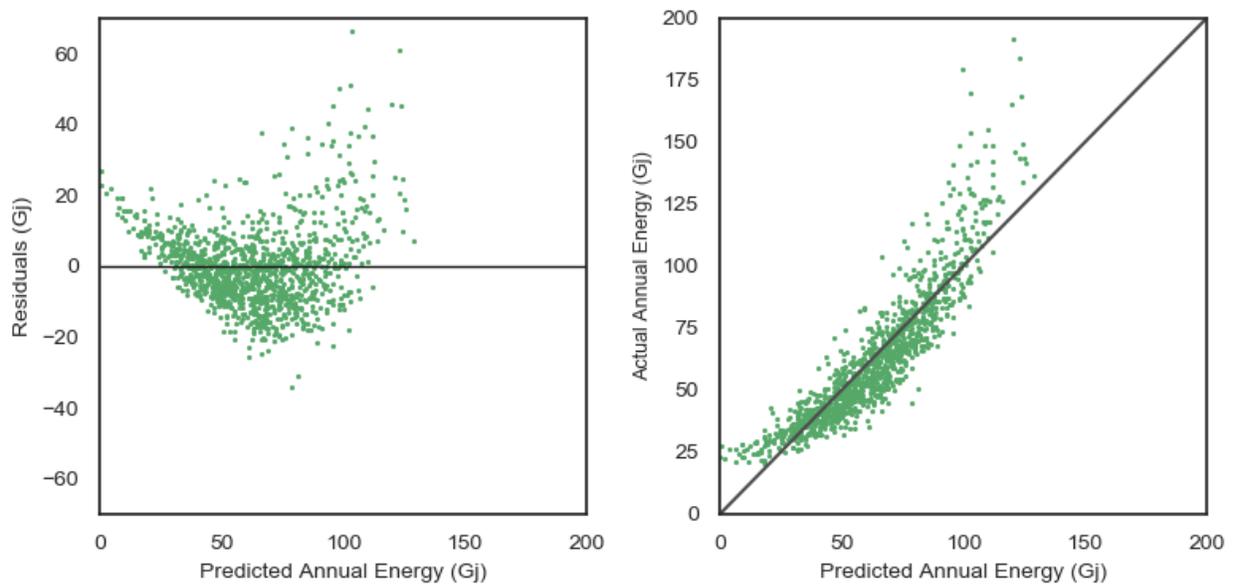
The residuals follow a U shape instead of a random distribution which indicates non-linear behaviour in the data. The predicted energy vs. actual energy graph shows that an untransformed linear model is a poor fit to the data. Actual energy refers to simulated energy calculated with EnergyPlus throughout this research.

Figure 35 shows the same graphs with the log-log transformed dataset, and these graphs confirm that the log-log model is a much better fit to the data.
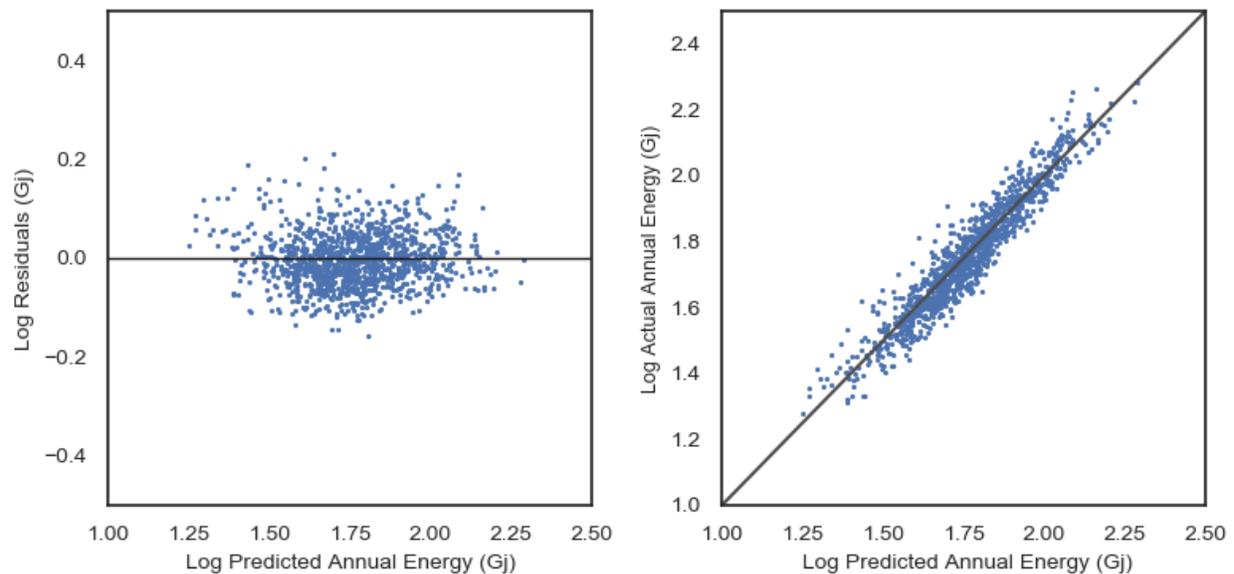
Figure 35 Log-Log transformation on the output and target variables, residuals and predicted vs. actual plots.

In order to interpret the results of the regression, the target variable had to be transformed back into units of Gj from units of log GJ by raising 10 to the exponent of the output value, this was referred to as back-transforming. The fit of the back-transformed model is shown in Figure 34. The fit of the back-transformed model does not appear to be as good as the fit of the log-transformed model for higher energy samples. The back-transformed model still fits the data

substantially better than the original non-transformed model. Barnes [30] reported similar findings.
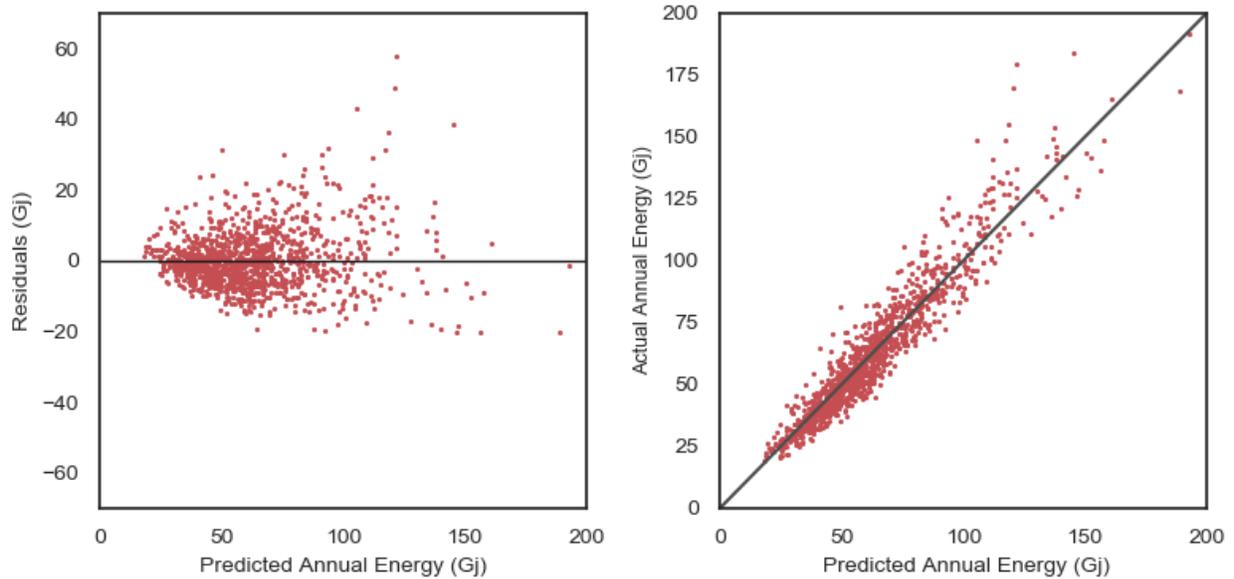
Figure 36 Back-transformed model showing residuals and predicted vs. actual plots in units of Gj

The distributions of feature distributions before and after logarithmic transformation were plotted for the combined model, in order to gain insight into the features prior to training. When a logarithmic transformation was applied to the original input features, the distributions of the base input features ended up skewed to the right. The original distribution of input features is shown in Figure 37. The log-transformed input feature distributions are shown in Figure 38.
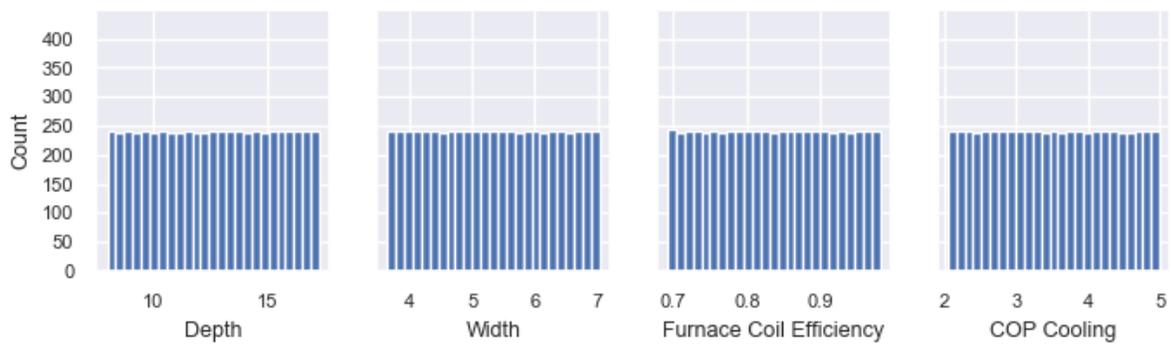
Figure 37 Input features before log transformation.
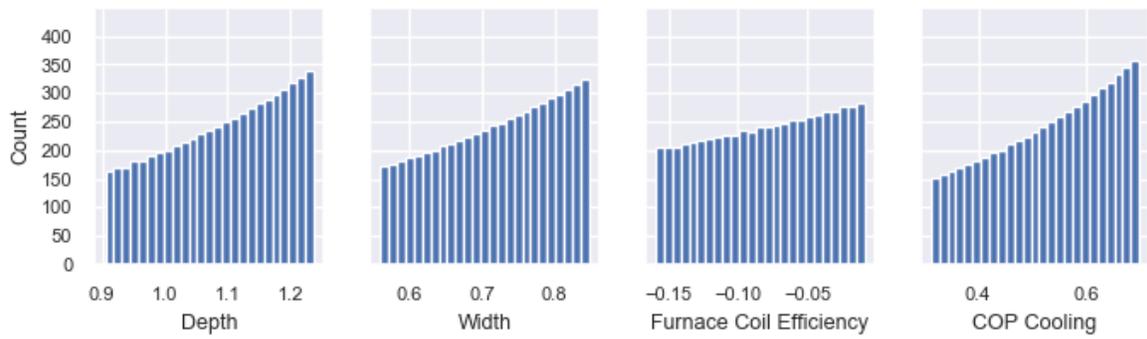
Figure 38 Input features after log transformation.

The target variable shows distribution that is slightly skewed to the left, applying a logarithmic transformation to the target variable creates a more normal distribution. These distributions are shown in Figure 39 and Figure 40
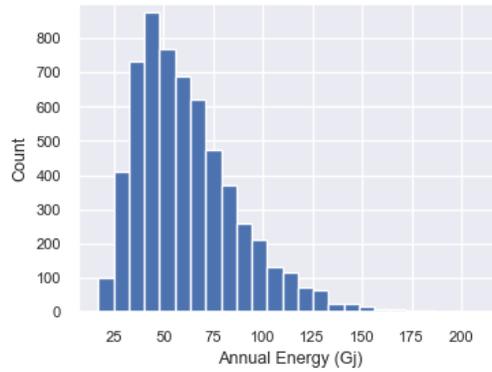
Figure 39 Energy use distribution before the transformation.
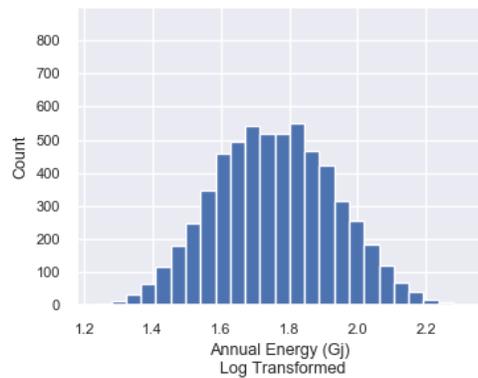
Figure 40 Logarithmic Transformation of Energy Use.

The geometry features that were generated by combining base input features, show

approximately normal distributions before and after logarithmic transformation. Some

examples of these features are shown in Figure 41.
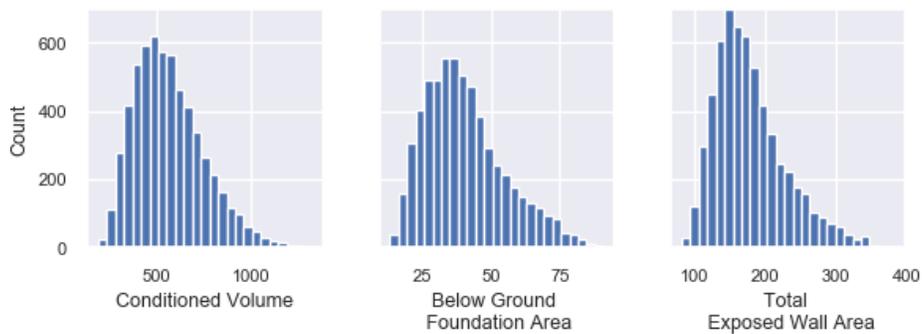
Figure 41 Generalized geometry feature distribution plots.

Features that had at large distributions of zero values were not included in the combined

archetype model due to their apparent non-normal distributions and semi-continuous numeric

properties. Examples of semi-continuous features are any features that include the right side of

the house. In the detached archetype, this wall will be defined. In the semi archetype, this side

of the house will be a shared wall. Some examples of these features are plotted below in Figure

42.
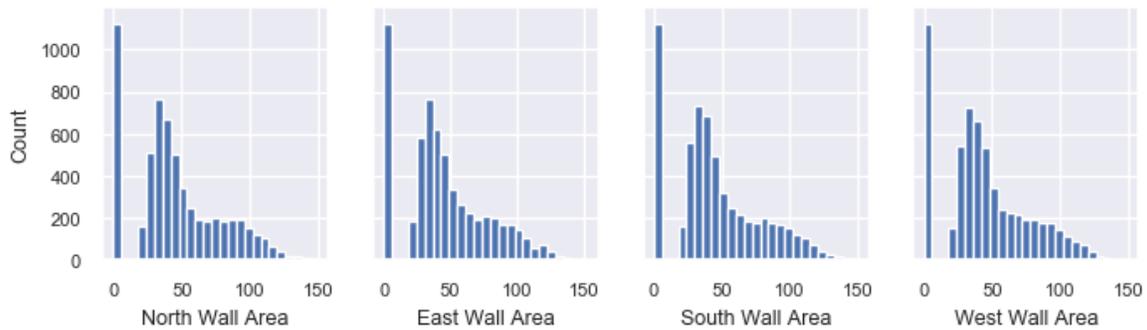
Figure 42 Distribution plots of semi-continuous features.

The effect of the log-transformation made the distributions significantly worse. This confirms

that these features should be excluded from the potential options during feature selection. The

log-transformed distributions of semi-continuous features are shown in Figure 43.
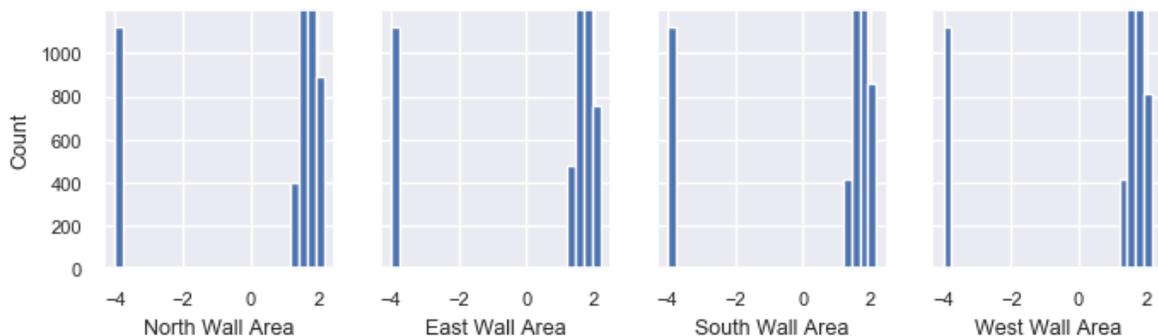
Figure 43 Log Transformed Distributions of semi-continuous features.

Combinations of these features were included in the model, the effect of adding opposite wall areas together removed the distribution at zero, the effect of adding semi-continuous and continuous features together was a distribution slightly skewed to the left. The untransformed distributions for these features are shown in Figure 44.
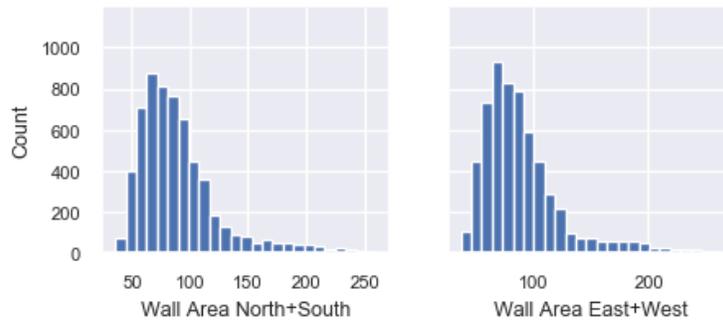
Figure 44 Distribution plots for combined semi-continuous features.

When a logarithmic transformation is applied to these new combined terms, the distribution shows a more normal distribution. Linear combination of semi-continuous and continuous features was an effective way to create normal distributions out of non-normal input features while still capturing the effect of that feature. The log-transformed distributions for these features are shown in Figure 45.
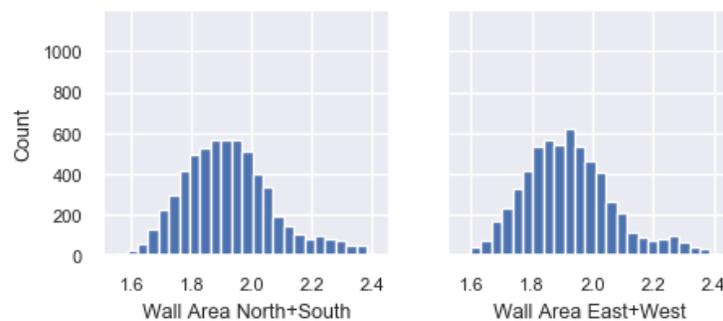
Figure 45 Distribution plots for combined semi-continuous features (log-transformed)

4.1.4   Stepwise Feature Selection

A preliminary univariate analysis was conducted to determine whether collinearity might be an issue in the model, and what features show collinearity with each other. The Pearson correlations between each base input feature are shown in Figure 46.
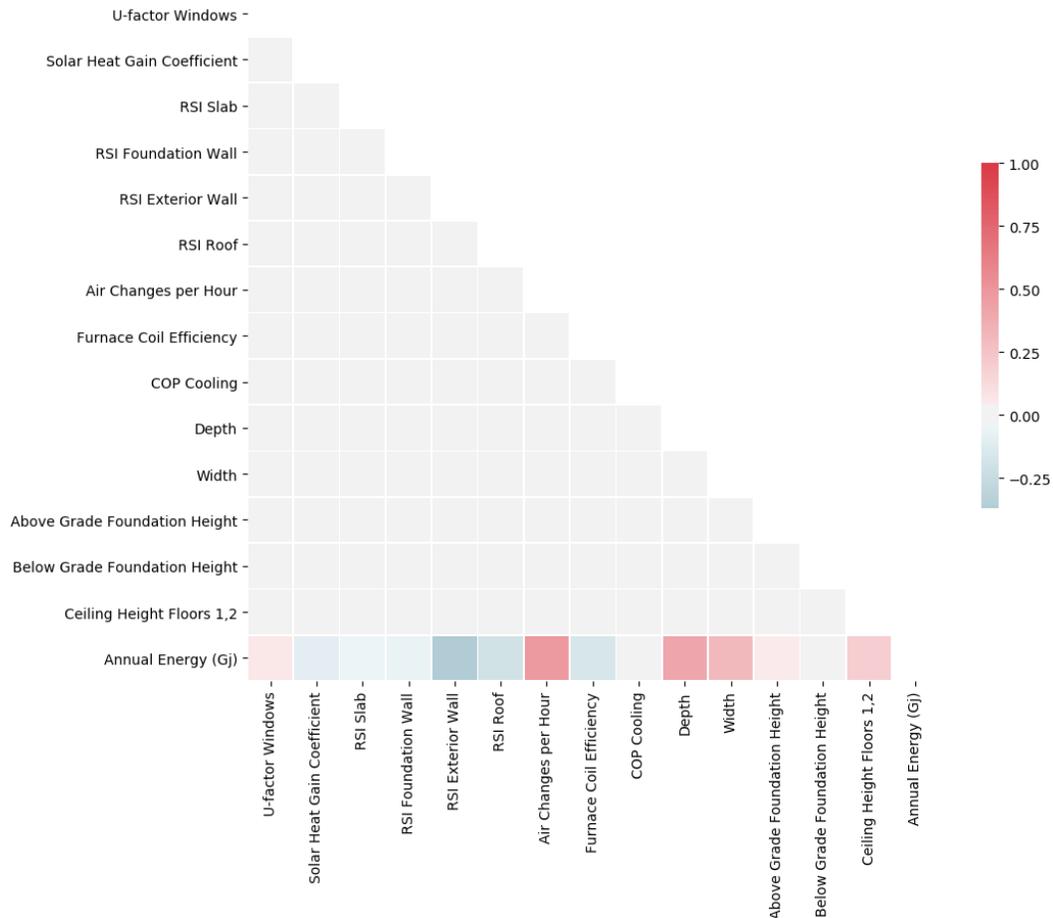
Figure 46 Pearson correlation original inputs and annual energy.

The base inputs are based on random values so there is no correlation, the heatmap verifies that there are no signs of collinearity. The univariate correlations to energy use are seen in the bottom row.

The Pearson correlation between each feature included in the set of feature options used for the combined model can be seen in Figure 47.
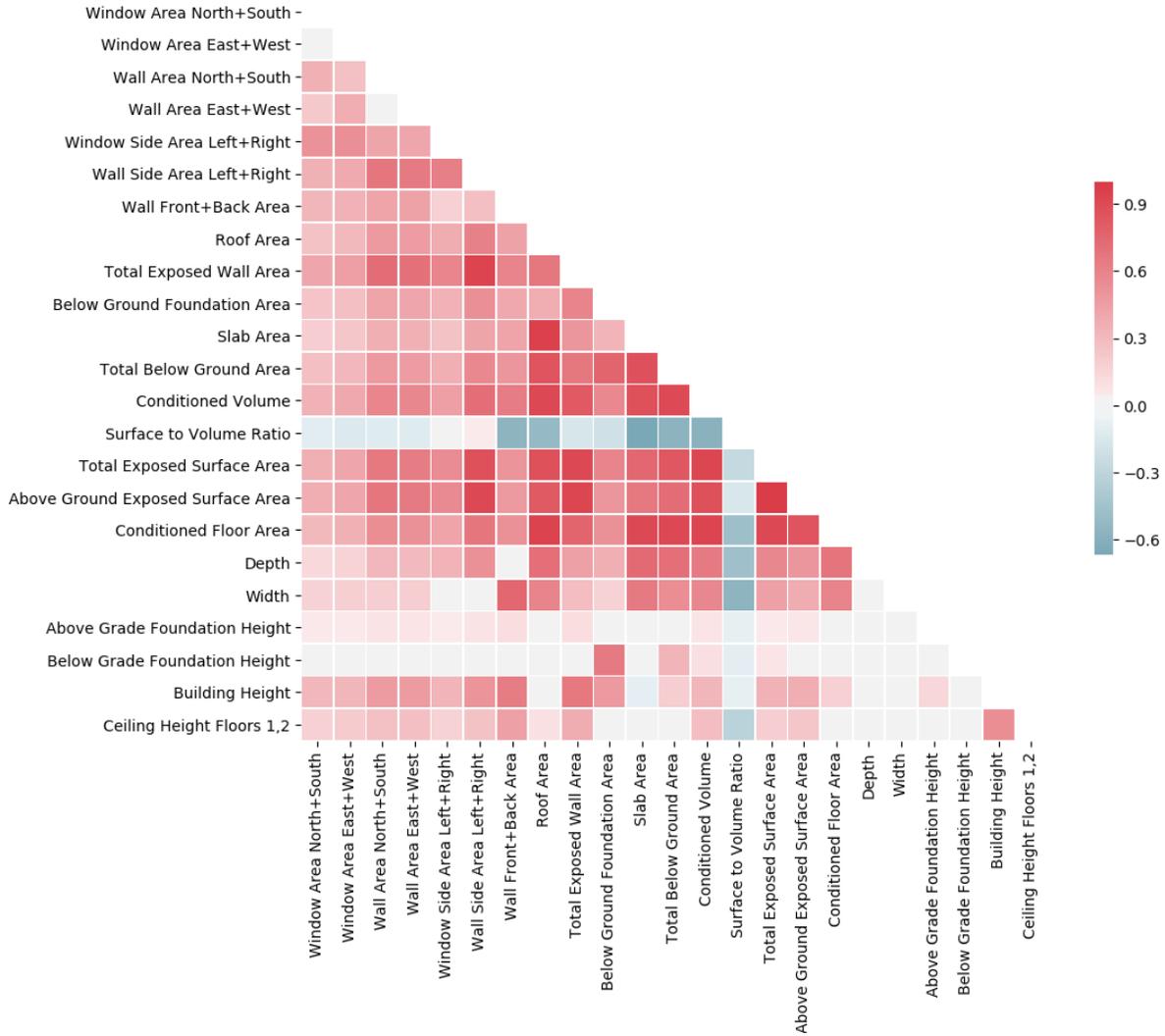
Figure 47 Pearson correlation all feature options for the combined dataset.

Figure 47 shows obvious signs of collinearity between features, many features share a Pearson correlation above 0.9. Although univariate correlations do not address multicollinearity, they help determine which feature combinations are causing collinearity. The variance inflation factor (VIF) is used to determine the amount of multicollinearity in the model. Features such as conditioned volume have very high VIF values, suggesting these features can be almost

perfectly predicted using a combination of other features in the model. Wall areas have the

highest VIF values of all the geometry features. Window areas have the lowest VIF values. The

VIF values for the set of feature options for the combined model are summarized in Table 14.

Table 14 VIF values for features options for the combined dataset.

Features:	VIF
Wall Area North+South	inf
Wall Area East+West	inf
Wall Side Area Left+Right	inf
Wall Front+Back Area	inf
Conditioned Volume	81581777013
Total Exposed Surface Area	54820662159
Surface to Volume Ratio	10653002993
Width	1029699
Conditioned Floor Area	493479
Depth	259922
Slab Area	99181
Building Height	85408
Above Ground Exposed Surface Area	2994
Total Below Ground Area	586
Total Exposed Wall Area	44
Below Ground Foundation Area	338
Ceiling Height Floors 1,2	286
Below Grade Foundation Height	206
Roof Area	141
Above Grade Foundation Height	30
Window Side Area Left+Right	4.7
Window Area East+West	3.1
Window Area North+South	2.9

During stepwise feature selection, the choice of variance inflation factor (VIF) Threshold

resulted in very different sets of features being selected. A value of 5 was chosen as the VIF

threshold based on research identified in Chapter 2. This implies that features can be included

that have an R^2 of up to 0.8 when regressed on all other features in the model. The effect of

varying the VIF threshold on the feature selection results for the combined dataset can be seen

below in Table 15. A higher VIF Threshold allows for more descriptive geometry features to be

included in the model. The accuracy of the model increases as the threshold for acceptable

collinearity in the model increases. It is interesting to note that as the VIF threshold increases, the accuracy score improves but quickly plateaus at a VIF threshold close to 5. Allowing more collinearity in the model past this point results in marginal improvements to model performance at the expense of a significant decrease in model interpretability. This trend was consistent across all house types. This confirms that choosing a VIF threshold of 5 or below results in a negligible performance sacrifice on all house types tested.

Table 15 Effect of VIF Threshold on stepwise feature selection.

Included Features	VIF Threshold in Stepwise Selection					
	1.5	2	5	10	100	inf
Window Area North+South			x		x	x
Window Area East+West						
Wall Area North+South			x	x	x	x
Wall Area East+West			x	x	x	x
Window Side Area Left+Right			x	x	x	x
Wall Side Area Left+Right					x	x
Wall Front+Back Area					x	x
Roof Area						
Total Exposed Wall Area						x
Below Ground Foundation Area			x	x	x	x
Slab Area						
Total Below Ground Area						
Conditioned Volume	x	x	x	x	x	x
Surface to Volume Ratio			x	x	x	x
Total Exposed Surface Area						x
Above Ground Exposed Surface Area						x
Conditioned Floor Area						x
Neighbouring House Distance	x	x	x	x	x	x
Number of Conditioned Floors	x	x	x	x	x	x
House Type	x	x	x	x	x	x
Grouped Orientation	x	x	x	x	x	x
Depth		x	x	x	x	x
Width						
Above Grade Foundation Height	x		x	x	x	x
Below Grade Foundation Height	x	x	x		x	x
Building Height						

Ceiling Height Floors 1,2						
Max Included Vif	1.49	2	4.99	5.29	61.28	6.85E+10
Validation MAPE	1.9032	1.8640	1.7546	1.7547	1.7460	1.7398
Validation R^2	0.9413	0.9435	0.9495	0.9494	0.9498	0.9502

4.1.5   Stepwise results: Individual Archetype models

Using stepwise feature selection, the optimal set of features was developed for each outer cross-validation fold. All geometry feature options were added to the base parameters iteratively, the feature that produced the greatest reduction to the validation error without increasing the VIF above a threshold of 5 was kept. This was repeated until the validation error stopped decreasing, or there were no more features left. The mean prediction metrics for each house type are reported below. The model predicted energy use between 7.03% to 7.54% of the simulated value.

Table 16 Results from training regressions on each archetype individually.

	Back Transformed		
	R^2	RMSE	MAPE
Small Two-storey Semi	0.93 ± 0.009	5.87 ± 0.46	7.54 ± 0.42
Large Two-storey Semi	0.935 ± 0.013	6.01 ± 0.68	7.03 ± 0.66
Three-storey Semi	0.939 ± 0.007	6.38 ± 0.28	7.35 ± 0.41
Detached	0.934 ± 0.009	7.63 ± 0.26	7.33 ± 0.67

The features selected using stepwise selection for each different split of the test, and training +validation sets are shown in the tables below and Figure 48. Depending on the archetype, different features were selected. Additionally, depending on the split of data being used, there

was inconsistency in the features selected. More training examples could potentially be used to increase the stability of these individual models. Figure 48 shows the number of times each feature is selected by each model type.
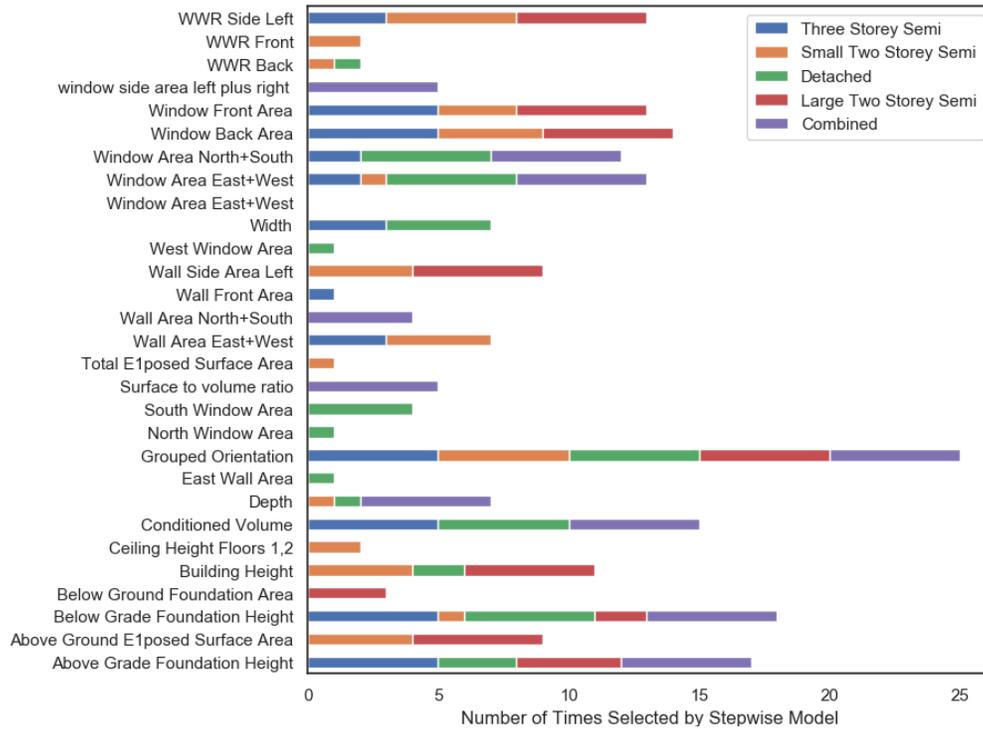
Figure 48 Number of times features are chosen by each model during stepwise selection.

The stepwise feature selection process is sensitive to small changes to accuracy when features are added. When not enough training examples were used the presence of a few outlying data points in the training or validation sets may be enough to cause the stepwise selection process to choose different features between folds. The features all have high multicollinearity so many of the features will have a similar effect on accuracy and a small change in data splits will result in different features being chosen. When more training examples were used the effect of outliers will be less significant, and the features selected in each fold will be more consistent.

The features selected in each outer cross-validation fold for each archetype are outlined in Table 17 to Table 20.

Table 17 Features selected by forward stepwise selection for the large two-storey semi in each fold.

Cross-validation Fold:	1	2	3	4	5
Above Grade Foundation Height		X	X	X	X
Above Ground Exposed Surface Area	X	X	X	X	X
Below Grade Foundation Height			X	X	
Below Ground Foundation Area	X	X			X
Building Height	X	X	X	X	X
WWR Side Left	X	X	X	X	X
Wall Side Area Left	X	X	X	X	X
Window Back Area	X	X	X	X	X
Window Front Area	X	X	X	X	X

Table 18 Features selected by forward stepwise selection for the small two-storey semi in each fold.

Cross-validation Fold:	1	2	3	4	5
Above Ground Exposed Surface Area	X	X		X	X
Below Grade Foundation Height			X		
Building Height	X	X		X	X
Depth	X				
Total Exposed Surface Area			X		
WWR Back	X				
WWR Front	X			X	
WWR Side Left	X	X	X	X	X
Wall Area East+West	X		X	X	X
Wall Side Area Left		X	X	X	X
Window Area East+West	X				
Window Area North+South					X
Window Back Area		X	X	X	X
Window Front Area		X	X		X

Table 19 Features selected by forward stepwise selection for the three-storey semi in each fold.

Cross-validation Fold:	1	2	3	4	5
Above Grade Foundation Height	X	X	X	X	X
Below Grade Foundation Height	X	X	X	X	X
Conditioned Volume	X	X	X	X	X
WWR Side Left		X	X		X
Wall Area East+West	X		X	X	
Wall Area North+South					X
Wall Front Area		X			
Width	X			X	X

	1	2	3	4	5
Window Area East+West	X		X		
Window Area North+South	X		X		
Window Back Area	X	X	X	X	X
Window Front Area	X	X	X	X	X

Table 20 Features selected by forward stepwise selection for the detached archetype in each fold.

Cross-validation Fold:	1	2	3	4	5
Above Grade Foundation Height	X		X	X	
Below Grade Foundation Height	X	X	X	X	X
Building Height	X	X			
Ceiling Height Floors 1,2			X	X	
Conditioned Volume	X	X	X	X	X
Depth			X		
East Wall Area					X
North Window Area				X	
South Window Area	X	X	X		X
WWR Back				X	
West Window Area		X			
Width	X	X		X	X
Window Area East+West	X	X	X	X	X
Window Area North+South	X	X	X	X	X

4.1.6   Stepwise results: Combined Archetypes

The combined dataset included 6000 samples from all four archetypes. The combined model was developed using the same forward stepwise methodology used for the individual archetype models. The combined model is used to determine whether similar accuracy can be reached with a surrogate model for multiple house types compared to a surrogate model for a single house type. The features selected in each fold of stepwise selection were almost identical across cross-validation splits, confirming that the combined model is much more stable than the individual models.

Using the stepwise methodology, the accuracy improved from an R^2 of 0.844 to 0.942 for the combined model. On average 30 features (counting each categorical option) were included in the final combined model by the feature selection methodology, this included 11 out of 45

potential geometry features. The feature selection methodology resulted in substantial

improvements to the accuracy and kept the VIF threshold below 5. Before feature selection, the

VIF was nearly infinite for some features. Figure 49 shows the fit of the model prior to stepwise

feature selection.
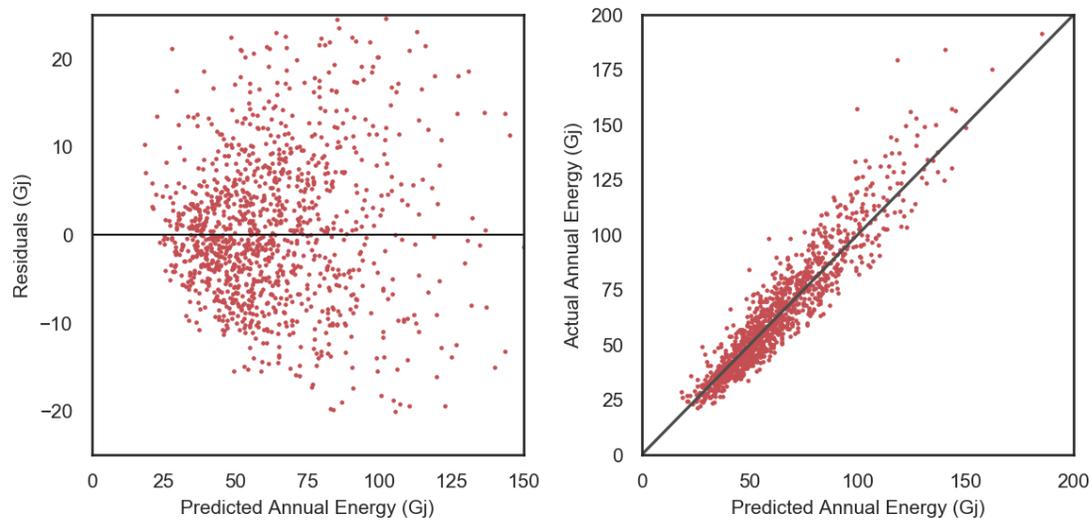
Figure 49 Residuals and predicted vs. actual plots before stepwise feature selection.

Figure 50 shows the fit of the model after the stepwise feature selection, the residuals are

substantially smaller and the predicted vs. actual plot shows a better fit as well.
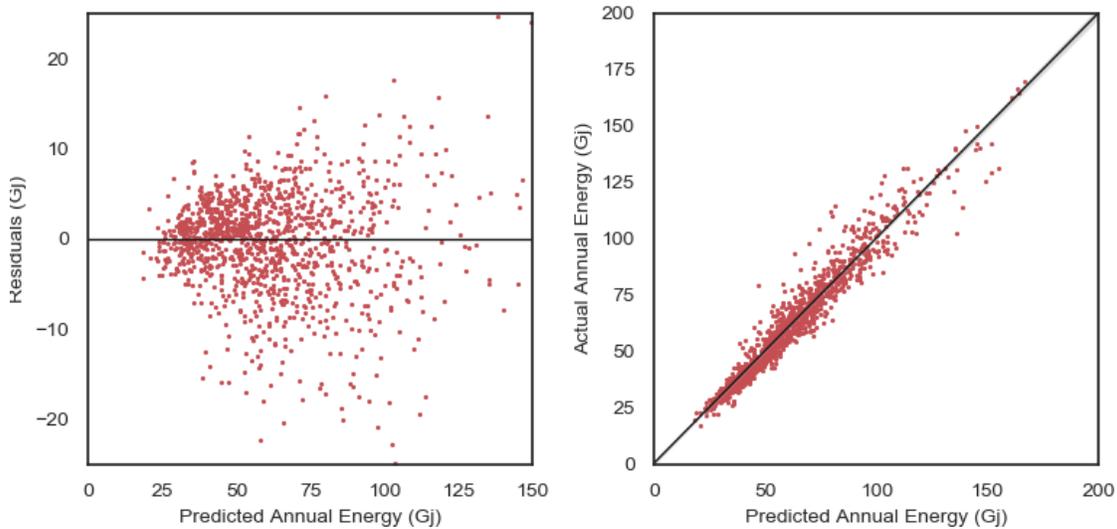
Figure 50 Residuals and predicted vs. actual plots after stepwise feature selection.

There is evidence of heteroscedasticity (the variance of the variance in the target variable increases) even after the stepwise selection process. The percent error is relatively constant throughout the model but the absolute error increases in relation to the amount of energy use. It is important to consider the maximum and minimum percent error when considering using a surrogate model. The data used to train the model was generated using latin hypercube sampling. For this reason, the data was not cleaned up and outlying data points were not removed from the set of training examples (energy models). The energy use for some training examples was poorly estimated by the surrogate model. The highest absolute percent error in predicting simulated energy use with the surrogate model was 37%, the lowest absolute percent error was 0.0082%. The MAPE of 7 for the combined model represents the average percent error, most of the individual training examples will have more or will have less error

than this average value. Future research should consider whether some outlying data points should be removed to improve the performance of the model.

The feature values for the 50 training examples which had the highest prediction error were studied to determine why the energy use for these houses could not be accurately represented by the surrogate model. One discernable pattern found in the houses with high percent error was that the R values for the roof, wall, foundation and slab had more variation. When the energy use was underestimated significantly the R value of the walls was substantially lower than average and the R value of the roof was higher than average. There is only a single control zone so it is very likely that this large range of R values between roof and wall for example will make it difficult for EnergyPlus to correctly size zones. This variation will create a larger variation in the energy performance of each zone and will make it harder to not undersize or oversize the amount of air delivered to each uncontrolled zone. The error may be coming from EnergyPlus overestimating or underestimating energy use, a more thorough investigation needs to be done to confirm. This high variation in R values does not represent a realistic combination of feature values that would be tested or found in an existing house. This is an important finding because the more realistic ranges of feature values are estimated more accurately by the surrogate model.

Many of the houses that were associated with higher prediction error also had unrealistic aspect ratios, such as a more square shaped house which would not occur in reality due to lot size, or a very long and narrow shape. When the extreme ends of the feature ranges were included in a single training example the error was often higher. Future work could examine the application of a set of rules which would remove unrealistic combinations of inputs from the

model. The accuracy could also be adjusted to account for these outlying examples. Figure 51 shows the effect of removing training examples with high absolute percent error. The training examples were sorted from highest percent error to lowest and removed one at a time and the MAPE was recalculated for each step.
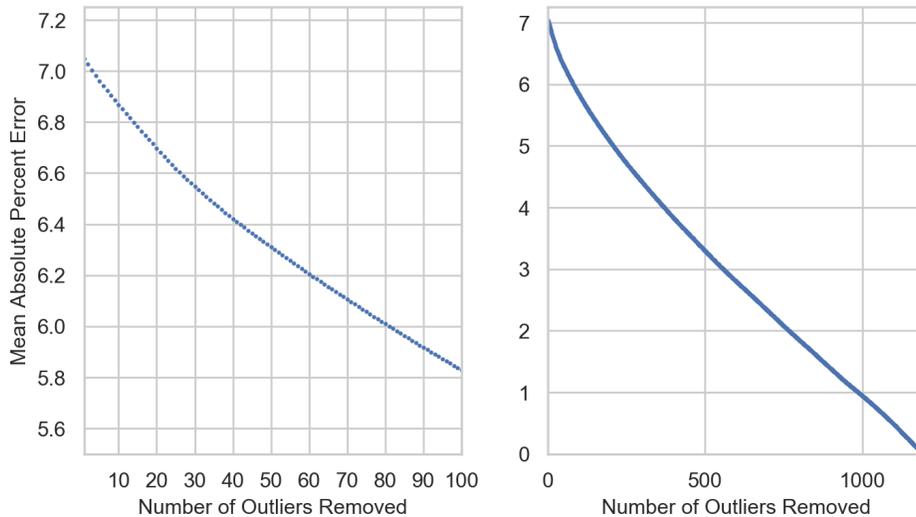
Figure 51 Mean absolute percent error as outlying training examples are removed.

The plots above show that the MAPE decreases quickly especially as the 40 worst examples are removed. Removing training examples from the model to improve accuracy should not be done unless substantially more work is done in order to develop a set of rules that define what combination of features are considered unrealistic. Removing training examples to improve the accuracy can result in an optimistic estimate of accuracy when used on real houses. Care must be taken to validate the reason for removing certain training examples if this approach is considered.

In order to verify that the combined model was predicting well for each archetype, the test set was subdivided into four archetype subsets. Each of these subsets was used as a test set and

the entire test set was also used as the test set. The prediction metrics reported in Table 21 show that the combined model predicted energy use with lower test error than the individual houses. This confirms that a single model may be used for at least four different house types. The larger number of samples likely contributed to the higher accuracy found with the combined model. The number of samples per house did not change, but the number of overall samples increased. This suggests that adding samples from different house types improves the prediction accuracy for individual houses. The accuracy of the stepwise combined model for each archetype and the full test set using the combined model is shown in Table 21.

Table 21 Combined stepwise model accuracy, predicting on subsets of the test set.

	Back Transformed		
	R^2	RMSE	MAPE
Full Test Set	0.942 ± 0.004	6.3 ± 0.16	7.03 ± 0.21
Three-Storey Semi	0.949 ± 0.01	5.83 ± 0.51	6.91 ± 0.51
Large Two-Storey Semi	0.936 ± 0.011	5.6 ± 0.69	7.22 ± 0.52
Small Two-Storey Semi	0.933 ± 0.016	6.01 ± 0.38	6.69 ± 0.47
Detached	0.936 ± 0.011	7.48 ± 0.58	7.26 ± 0.44

The features selected during each fold of the forward stepwise selection methodology for the combined model are shown in Table 22. The model selects the same features in nearly every fold of the outer cross-validation loop, the combined model with 6000 total samples is substantially more stable than the individual models generated with 1500 total samples.

Table 22 Features selected in each fold by the combined model

Cross-validation Fold: combined	1	2	3	4	5
Above Grade Foundation Height	X	X	X	X	X
Below Ground Foundation Area	X	X	X	X	X
Conditioned Volume	X	X	X	X	X
Depth	X	X	X	X	X
Surface to Volume Ratio	X	X	X	X	X
Wall Area North+South	X	X	X	X	
Window Area East+West	X	X	X	X	X
Window Area North+South	X	X	X	X	X
Window Side Area Left+Right	X	X	X	X	X
Grouped Orientation	X	X	X	X	X
Number of Conditioned Floors	X	X	X	X	X
House Type	X	X	X	X	X
Neighbouring House Distance	X	X	X	X	X

The standardized regression coefficients generated by the regression model are shown in Table 23, sorted from largest to smallest. Note that a smaller window U-factor range than intended was used so in reality the window related features might be slightly larger.

Table 23 Standardized regression coefficients generated using feature selection methodology

stepwise	Mean	Sd
Conditioned Volume	0.1599	0.00144
Air Changes per Hour	0.1094	0.00044
Ventilation/area	0.0456	0.00077
Surface to Volume Ratio	0.0226	0.00204
Neighbour:yes	0.0191	0.00034
3 Storeys	0.0162	0.00138
Window Side Area Left+Right	0.0139	0.00041
U-factor Windows	0.0119	0.00024
Depth	0.0084	0.00029
West	0.0066	0.00065
North	-0.0011	0.00112
Wall Area North+South	-0.0016	0.00034
Above Grade Foundation Height	-0.0018	0.00020
COP Cooling	-0.0021	0.00025
Fluorescent	-0.0044	0.00074
LED	-0.0053	0.00034
RSI Slab	-0.0116	0.00026
RSI Foundation Wall	-0.0123	0.00023
East	-0.0130	0.00116
Window Area East+West	-0.0180	0.00053
Window Area North+South	-0.0183	0.00042
Solar Heat Gain Coefficient	-0.0191	0.00011
Below Ground Foundation Area	-0.0260	0.00041
Furnace Coil Efficiency	-0.0298	0.00019
RSI Roof	-0.0405	0.00050
60% HRV	-0.0515	0.00046
RSI Exterior Wall	-0.0718	0.00033
85% HRV	-0.0761	0.00107
90% HRV	-0.0818	0.00078
Detached	-0.0831	0.00327
Intercept:	1.8175	0.00104

The coefficients are shown in descending order. The inputs were standardized during

preprocessing, so all features are on the same scale. A unit change in the input corresponds to a

unit change to the output. Negative values correspond to decreasing the energy use when the

input value increases. The importance of each feature can be seen here. Conditioned volume

has the largest positive absolute value, followed by air changes per hour at 50Pa and normalized mechanical ventilation rate. The efficiency of the heat recovery system was considered quite important because the ranges used cover baseline to passive house. The efficiency of heat recovery will be important when lower energy intensity and air changes per hour are achieved. Changing the wall insulation RSI value has a greater impact than changing the RSI of the roof, likely due to the larger surface area

4.1.7   Stepwise Results: Unseen Archetype

In order to investigate the ability of a surrogate model to predict energy use for a house type not included in the training set, a series of models were developed. Each archetype was held out from the dataset to be used as a test set, and a new model was trained on the remaining three archetypes. The results of these tests give a range of values that could be expected if a surrogate model was to be used to model any archetype within the century single-family category. In these tests, only three different house types are used for each test, and only 4500 samples total. If all four archetypes were used to train these models, they would be even more comprehensive and be able to predict on a different house type more accurately. This error estimate can also be considered quite conservative because the variation between archetypes in this model quite large and considers different numbers of floors and semi and detached houses. The scope of a single model can also be improved further by adding additional training examples of different geometries to cover more variation in building form.

These models were developed using the stepwise feature selection methodology, and nested cross-validation was used to determine the mean prediction error for each house and SD. The holdout test scores are summarized in Table 24 and Figure 52.

Table 24 Summary of test results using an unseen archetype as a hold out evaluation set.

Archetype				Back Transformed Scores		
Detached	Three-storey Semi	Small Two-storey Semi	Large Two-storey Semi	R²	RMSE	MAPE
Holdout Set	Train Set	Train Set	Train Set	0.760	14.608	14.431
Train Set	**Holdout Set**	Train Set	Train Set	0.931	6.801	7.796
Train Set	Train Set	**Holdout Set**	Train Set	0.906	7.242	8.064
Train Set	Train Set	Train Set	**Holdout Set**	0.816	9.569	14.599

4.1.7.1 Detached Holdout
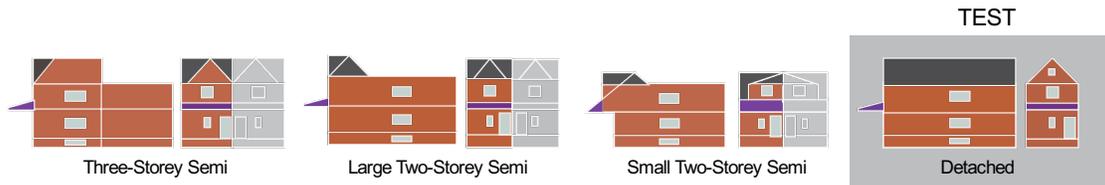
When using a model trained exclusively on semi archetypes to predict the energy use of a detached archetype, the average MAPE was 14.43%. The quality of fit for this model on the detached archetype is shown in Figure 52. The residuals show a noticeable non-normal trend that suggests the model is underestimating energy use for the detached houses when it is only trained on semi detached houses. This was expected because a detached house would have a larger exposed surface area, and the model was not trained on any houses with shared walls.
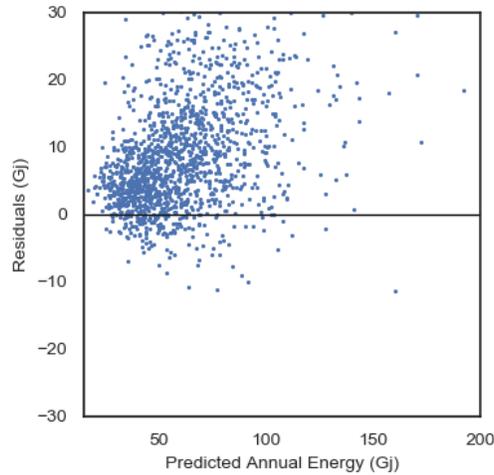
Figure 52 Detached holdout test residual plot.

4.1.7.2   Three-storey Semi Holdout
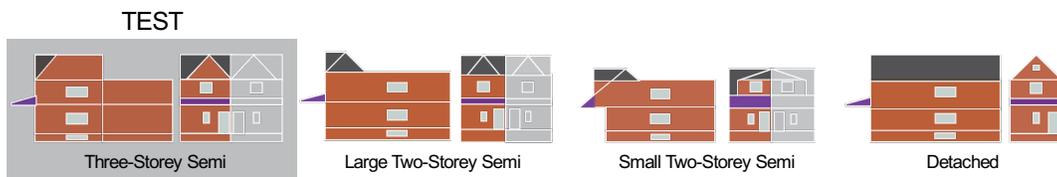
Energy use for the Three-Storey Semi archetype was predicted using a model trained on the

remaining models, this model predicted with an average MAPE of 7.79%. The effect of the

house type parameter was investigated to see if this was a crucial feature in determining

energy use. It was found that the house type feature substantially decreased accuracy when

included in the model. The reported accuracies in Table 24 are for the models without house

type included.
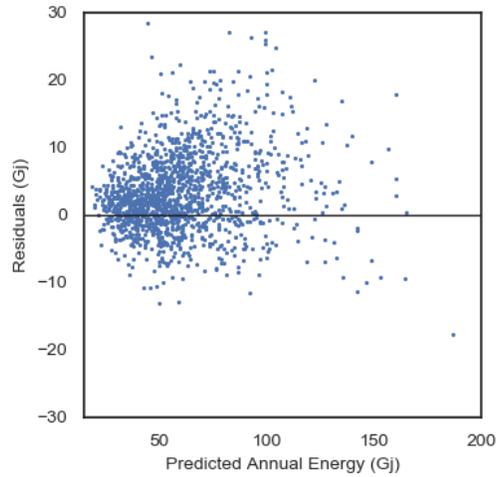
Figure 53 Three-storey semi hold out test (with "House Type" residual plot).

4.1.7.3 Small Two-Storey Semi Holdout

TEST

| Three-Storey Semi | Large Two-Storey Semi | Small Two-Storey Semi | Detached |
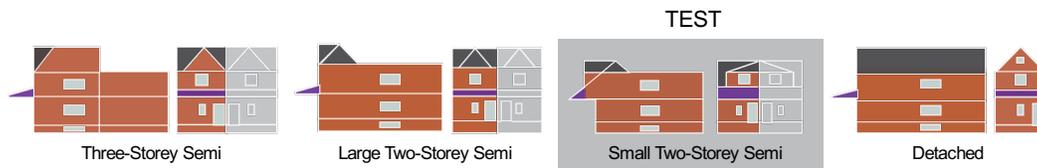
In Figure 54, the quality of fit can be seen for the model tested on the small two-storey semi.

The fit of this model can be compared to Figure 55, the residuals are more centred on 0 and the

predicted vs. actual plot shows a tighter fit and less evidence of non-normality.

114
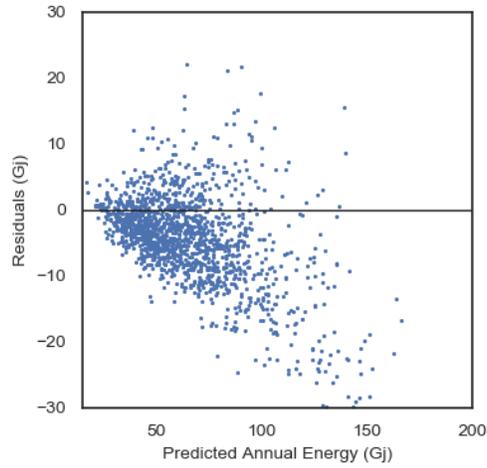
Figure 54 Small two-storey semi house type hold out test (with "House Type") residual plot.
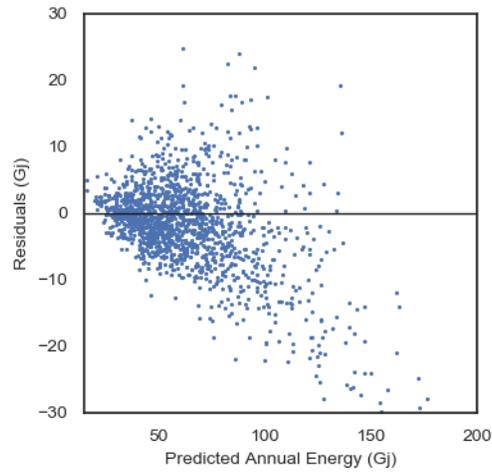
Figure 55 Small two-storey semi holdout test (without "House Type") residual plot.

4.1.7.4   Large Two-Storey Semi Holdout

TEST

Three-Storey Semi   Large Two-Storey Semi   Small Two-Storey Semi   Detached
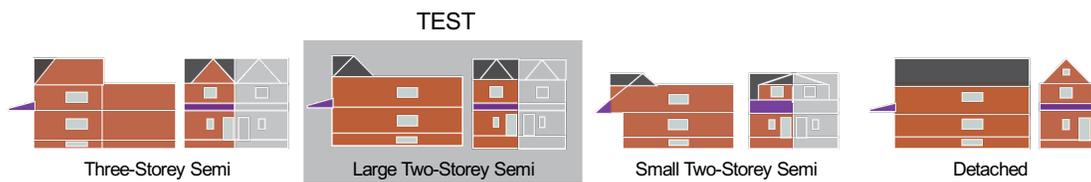
The accuracy when predicting using the large two-storey semi archetype as a holdout set was the lowest of all archetypes tested. This is surprising because it was assumed that the energy use for the detached house type would be the most difficult to predict because it is the only non-semi archetype. The two-storey semi geometry was the most simple, and potentially the exposed surface area to volume ratio would have been lowest in this house. This would explain why the results suggest an over-estimation of energy use for this house. Training the models using only archetypes with high energy use intensity because of higher surface to volume ratio will result in a model that places more significant weight to feature coefficients such as floor area or volume. The coefficient estimates would not be as representative of the more simple geometry seen in the large two-storey semi. The "House Type" feature made the predictions worse in this model as well. The comparison can be seen in Figure 56 and Figure 57.
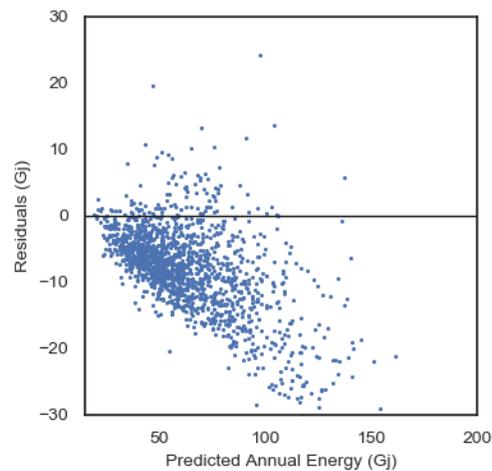
Figure 56 Large two-storey semi holdout test (without"House Type") residual plot
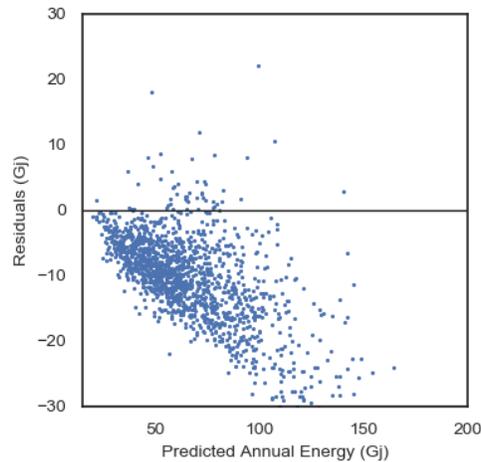
Figure 57 Large two-storey semi (with "House Type") residual plot

The average percent error drops from 20.55% to 14.6 when "House Type" is not included. This phenomenon is likely due to relationships between house type and other geometry features. The change in the exposed surface area associated with the difference between semi and detached houses is captured by other features already, which means the "house type" feature may be capturing other characteristics of the house unrelated to the party wall area.

4.1.8 Stepwise results: Sample Size Analysis

In order to ensure that the training sets were sufficiently large, and to further explore the stepwise regression model, a regression was run with a decreasing number of training samples (energy model simulations) and plotted against test error. As the training sample size increases, the testing and training scores tend to converge towards each other. If the number of training examples is large enough, the training and testing scores should converge. An important finding from the sample size analysis is that the number of training samples required

117

for the individual archetypes does not vary substantially from the number of samples required

for the combined model.

The combined archetypes model included 6000 samples made up of 1500 from each Archetype.

A test, train, and validation split of 25%,8% and 72% were used in the model. The training +

validation set contained 4800 samples for the combined model and 1200 for the individual

models. Figure 58 and Figure 59 show the results of the sample size analysis for the combined

archetype model. The Testing and Training scores appear to converge and plateau at a or

before a training size of only 2000. The number of total training examples could be reduced to

approximately 2000 without sacrificing accuracy based on these findings. Results suggest that

suggest the number of training examples from each archetype can be reduced to approximately

500 without sacrificing accuracy.
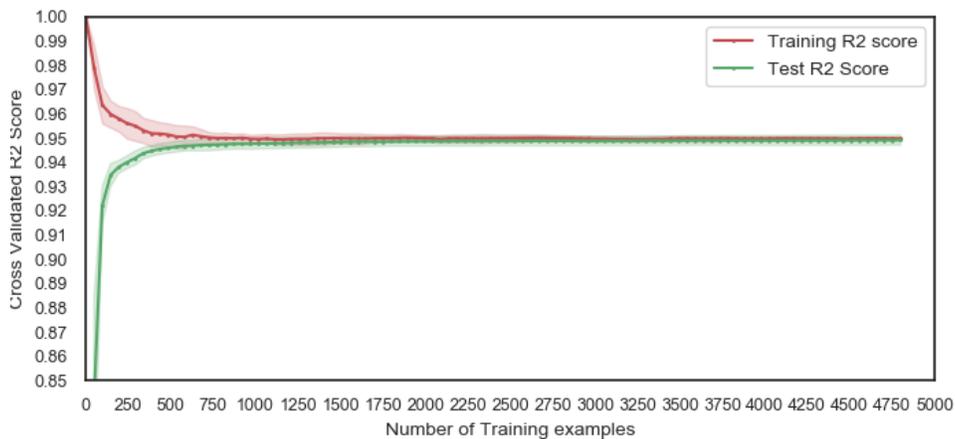
Figure 58 Combined model sample size analysis 0-5000

Figure 59 was included to allow comparison of the combined model to the individual models on

the same scale.
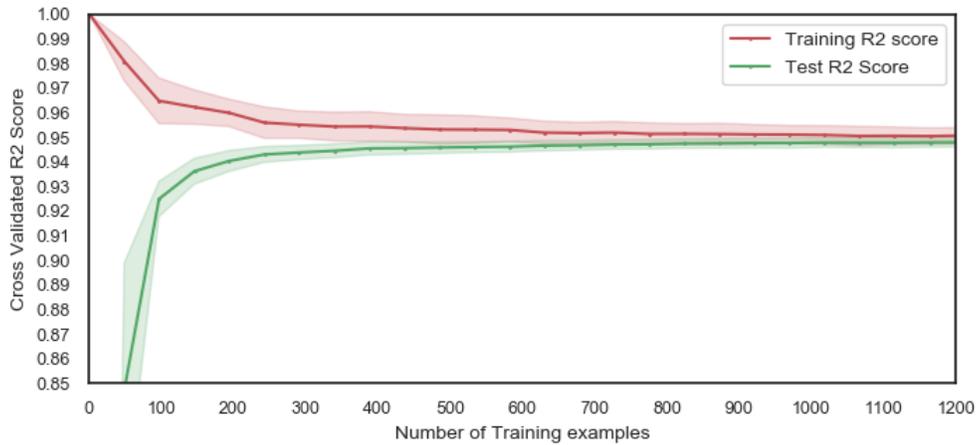
Figure 59 Combined model sample size analysis 0-1200

The convergence of the testing and training scores also signifies that the variance in the model is low, this explains why regularization with elastic net only tended to increase error as the tuning parameter value increased, if there was a lot of variance in the model elastic net can reduce test error by reducing variance. The standard deviation of the test scores is also small and doesn't change considerably; this reflects the stability of the model during feature selection. The standard deviation is represented by the shaded regions in the figures below.

Using 1200 training examples to build a single archetype model, the test and training RMSE show approximately the same convergence as the combined archetype model with the same number of samples. This analysis shows that by combining four archetypes into a single model, the simulation time per house to build the database of energy models can be reduced by four times. More research is required, but these findings suggest that as the number of archetypes described by a single model increases, the number of training examples required does not increase linearly. As more archetypes are included in a model, the number of training examples

119

per archetype will decrease. The sample size analysis for the individual archetype models can be seen in Figure 60 to Figure 63.
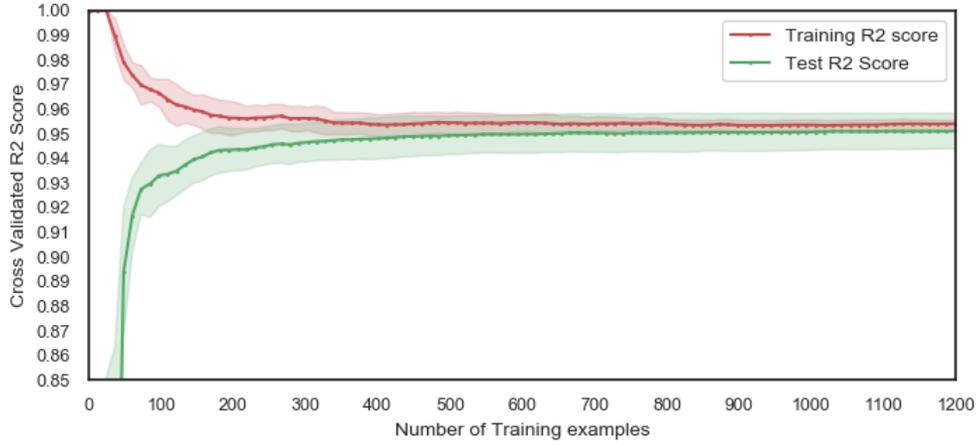
Figure 60 Detached sample size analysis
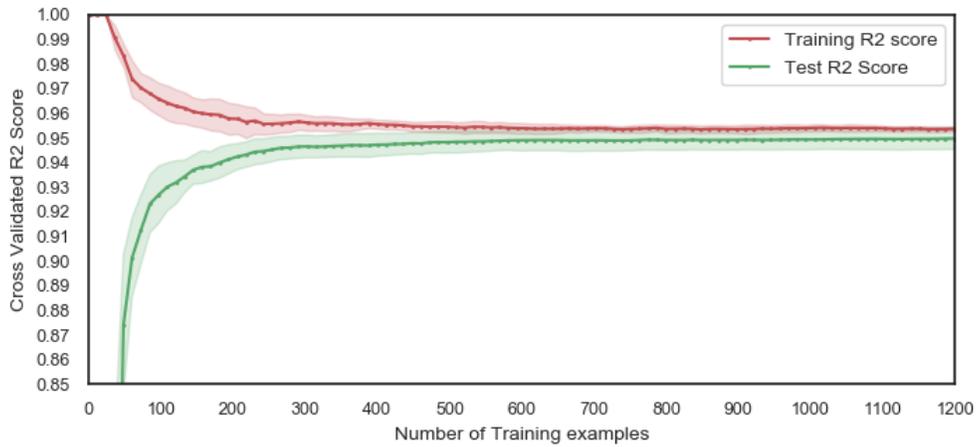
Figure 61 Three-storey semi sample size analysis
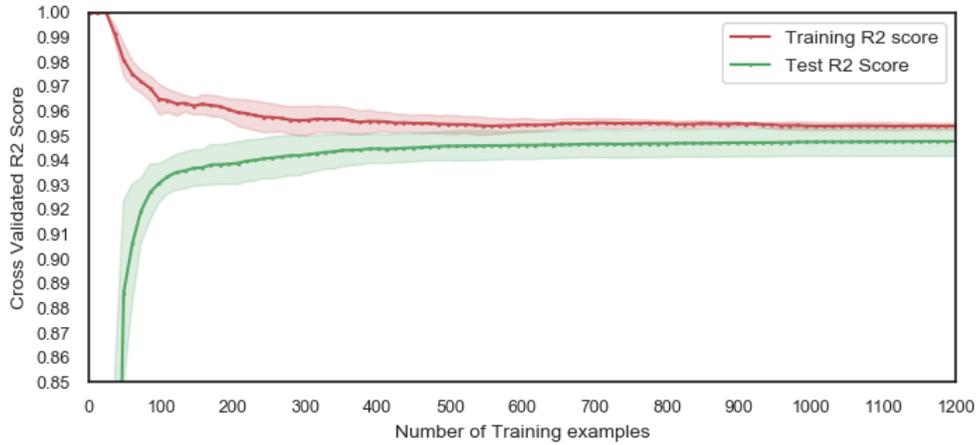
Figure 62 Large two-storey semi sample size analysis
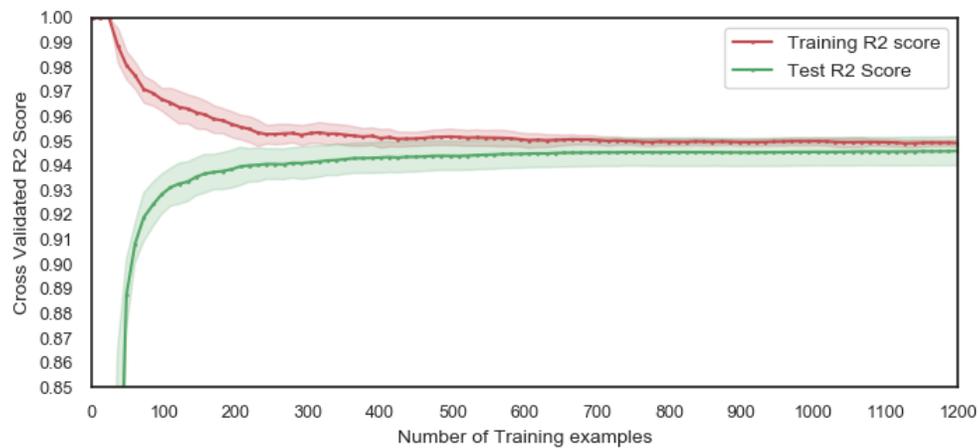
Figure 63 Small two-storey semi sample size analysis

The sample size analysis for the individual archetypes also explain the slightly lower accuracy in the individual archetype models compared to the combined model, the testing and training results appear to be still converging slightly past the 1200 sample point. If the test and training scores are still converging, the test score will still increase slightly if more samples were included.

The instability identified in the nested cross-validation results of the stepwise selection for individual archetypes can also be explained by the failure of the training and test errors to converge fully. The larger standard deviations in the prediction accuracy also suggest instability in the individual models compared to the combined model. Using more samples rather than less is the most conservative approach if computational time allows. A larger number of training examples may not improve accuracy substantially but will likely increase the stability of the model and its coefficients for the individual model.

4.1.9   Embedded Feature Selection Combined dataset model

In order to compare between elastic net and stepwise feature selection, an additional model was developed for the combined dataset. The process of developing and running a stepwise algorithm can be very time consuming; in addition, the optimal combination of features may not be selected depending on the order that features are added.

Elastic net regression was tested in this section as an alternative to the stepwise selection. The same set of 50 collinear feature options that were used for the stepwise regression were used as the starting feature set for the elastic net regression. An L1 value of 0.65 was found to provide the most desirable combination of lasso and ridge properties. A range of 1500 l values from 0.00001 to 0.06 were used to develop a set of models that were analyzed.

Figure 64 shows the RMSE score resulting from various l values. The RMSE value is plotted against the number of non-zero coefficients in the model.
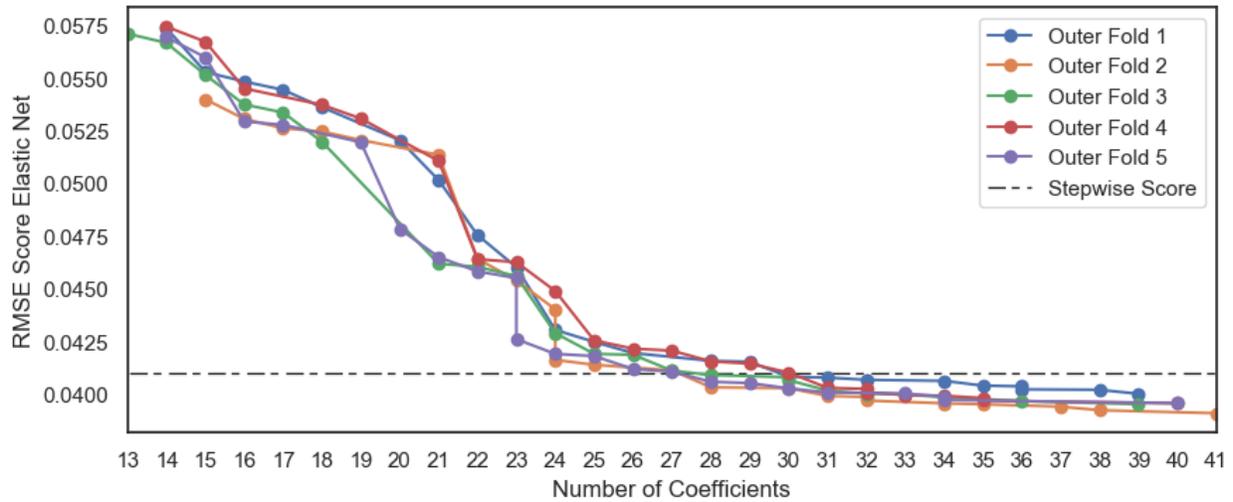
Figure 64 RMSE compared to number of coefficients.

Figure 65 shows a zoomed-in snapshot of the results shown in the previous figure. The RMSE

achieved by the stepwise is plotted as a horizontal dashed line.
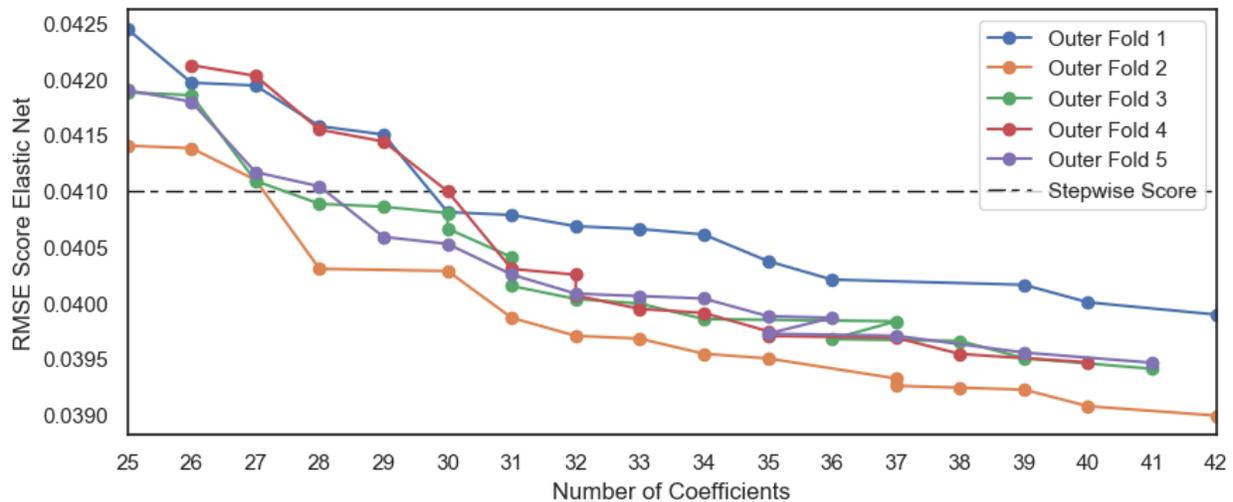
Figure 65 RMSE compared to the number of coefficients, zoomed in.

Elastic net can produce a model with lower error than the stepwise model if more features are

used, the elastic net and stepwise models appear to include approximately the same number of

features for a model with similar accuracy. The error starts rapidly increasing when there are approximately 30 features remaining in the model.

Most of the features removed from the elastic net model at small l values include categorical options. A large number of categorical options are removed early in part because all options are included in elastic net regression. In unregularized regression, one categorical option is dropped during the pre-processing phase to eliminate collinearity between options.

Models with a validation accuracy at or above the stepwise model accuracy(shown by the dashed line), included between 28 to 30 features depending on the split being trained on. The stepwise selection methodology selected a model with 30 features.

The average estimated coefficient values are plotted against tuning parameter values in Figure 66. Many coefficients are very quickly are reduced to zero or removed from the model at a very small tuning parameter.
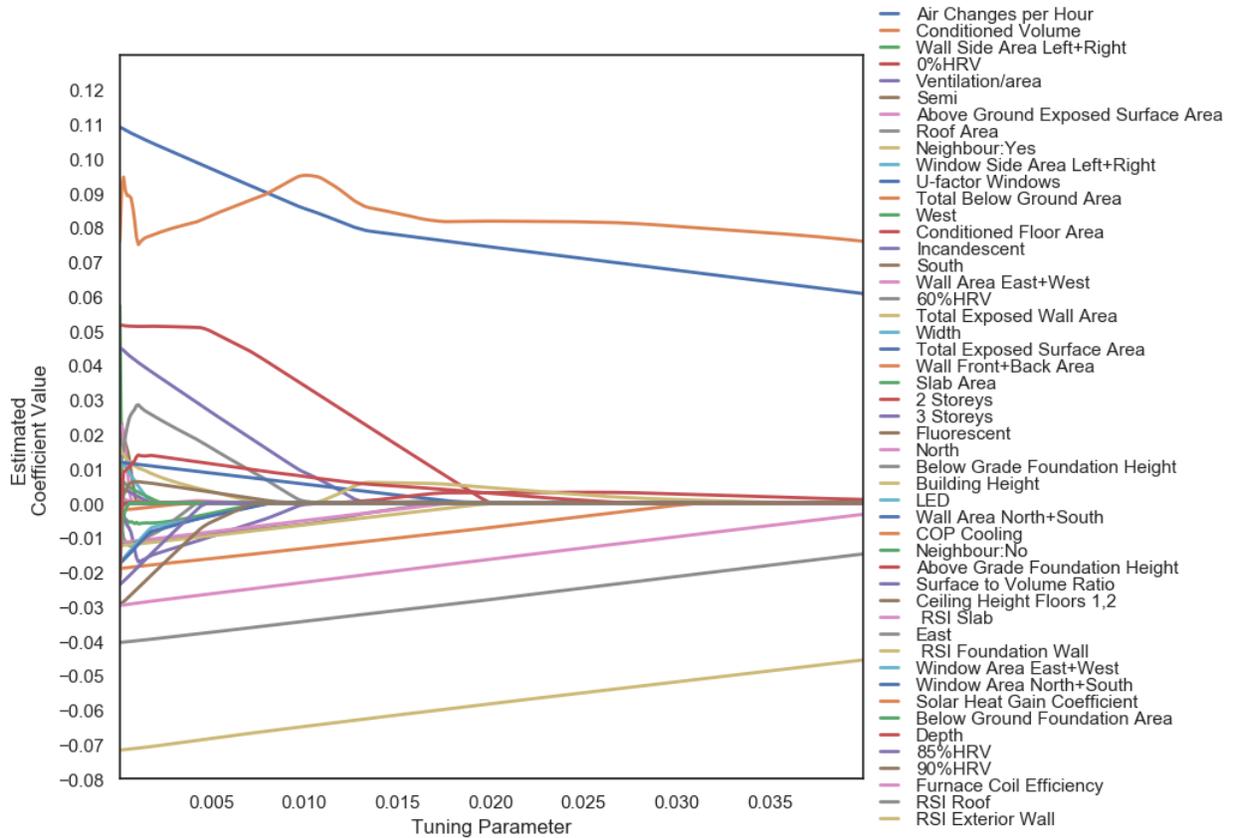
Figure 66 Coefficients plotted against tuning parameter l.

The erratic behaviour of estimated coefficients at a small tuning parameter value is due to

collinearity, which causes coefficients with a high variance to be quickly driven to 0. Some

features are reduced, and others are increased to compensate. The coefficients appear to

stabilize at a certain point in the model and decrease at a steady rate. As features describing

geometry are shrunk towards 0, the volume becomes a more critical predictive feature and

eventually surpasses air changes per hour. Generalized full building geometry features are

shrunk to zero due to very high collinearity; these features are all very close to linear

combinations of each other. Some categorical features that are quickly dropped from the

model include light type, number of conditioned floors, house type categories. In a separate analysis that was not described in this research, light type was found to have a high correlation to electricity end-use in the archetypes. The reduction in heat gain from more efficient light types resulted in an increase in heating demand. The result was a marginal increase in total energy use

The average tuning parameter selected is plotted as a vertical line in Figure 67 against the average coefficient values across five splits, this line was plotted on a smaller range of tuning parameters to illustrate the behaviour before and after the tuning parameter value. There appears to be a point on this plot where the estimated coefficient values begin to stabilize. When some coefficients shrink much quicker than others or change signs, it is a marker of collinearity in the model. Coefficients with higher variance are penalized more than other coefficients. Plotting coefficient estimates compared to tuning parameters could be an effective way to diagnose collinearity in a model and determine what tuning parameter will provide interpretable coefficients.
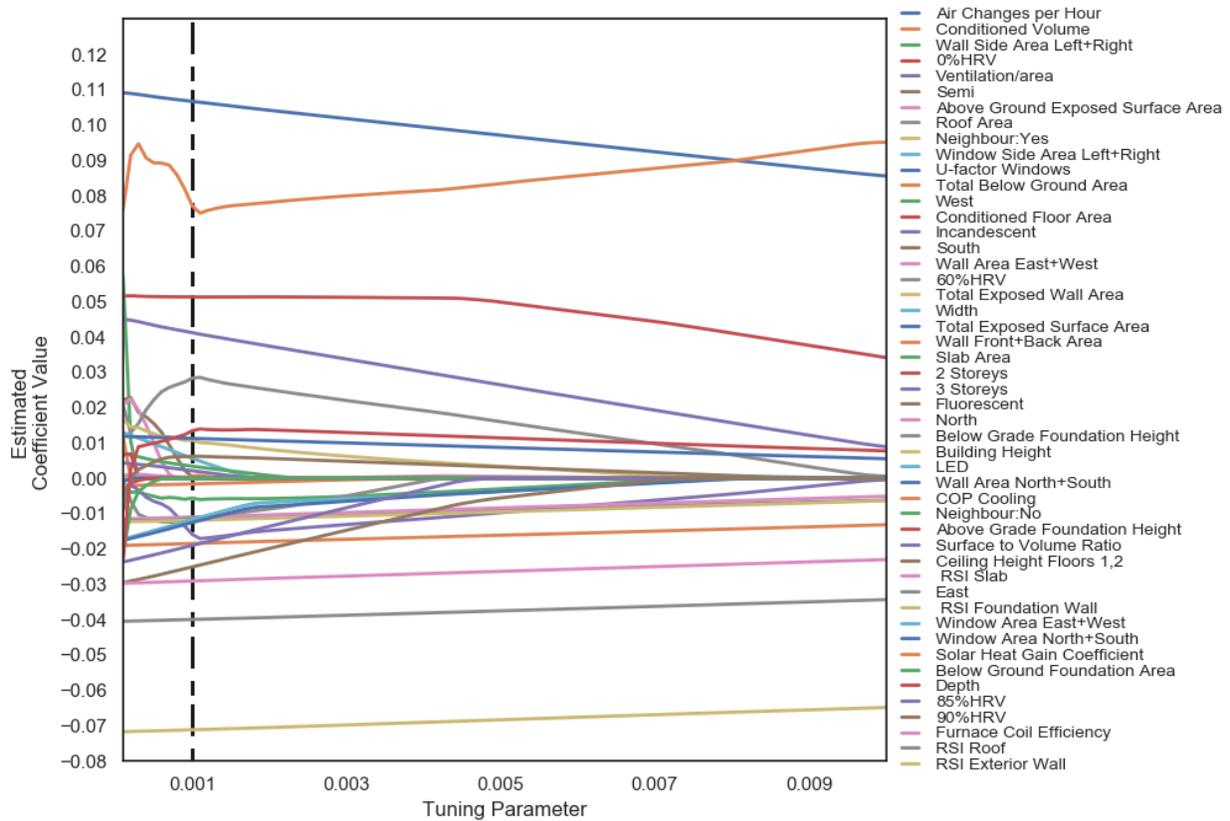
Figure 67 Coefficients plotted against tuning parameter l.

The estimated coefficients for the combined archetype model are shown below in table 7 ordered by descending value. These coefficients were calculated using the average of the tuning parameters from each split of the dataset, the full training/validation set. Insights on the importance of each feature can be taken from the sign and the magnitude of these coefficient values. Window features such as window area and window U-factor should be slightly more important than shown if the intended U-factor range was used. A smaller than intended range of U-factors was used. This research is intended as a methodology, so this unrealistic range will not affect the final conclusions. If this model was to be used in the future a more realistic U-factor range should be used for the windows.

	Mean	Sd
Air Changes per Hour	0.1066	0.0003
Conditioned Volume	0.0771	0.0030
0% HRV	0.0513	0.0004
Ventilation/area	0.0412	0.0005
Roof Area	0.0284	0.0018
Depth	0.0135	0.0005
U-factor Windows	0.0113	0.0003
Neighour:Yes	0.0108	0.0002
Ceiling Height Floors 1,2	0.0062	0.0004
Window Side Area Left+Right	0.0056	0.0006
West	0.0034	0.0003
Incandescent	0.0021	0.0005
Wall Front+Back Area	0.0006	0.0007
Semi	0.0004	0.0003
Wall Area East+West	0.0003	0.0003
COP Cooling	-0.0016	0.0003
Detached	-0.0028	0.0020
Neighbour:No	-0.0056	0.0001
RSI Slab	-0.0110	0.0003
East	-0.0113	0.0005
Window Area East+West	-0.0117	0.0007
Below Grade Foundation Height	-0.0117	0.0006
RSI Foundation Wall	-0.0119	0.0002
Window Area North+South	-0.0125	0.0004
Surface to Volume Ratio	-0.0159	0.0021
Solar Heat Gain Coefficient	-0.0184	0.0001
85% HRV	-0.0190	0.0011
90 % HRV	-0.0251	0.0010
Furnace Coil Efficiency	-0.0290	0.0002
RSI Roof	-0.0400	0.0007
RSI Exterior Wall	-0.0712	0.0005
Intercept	1.7543	0.0022

Table 25 Estimated Coefficients after elastic net regularization

The final model, with an average of 30 features, predicted energy use with 7.28% MAPE. The

model was also tested on subsets of the test set attributed to each house type. The model

predicted energy use between 6.87% MAPE for the Small Two-Storey Semi and 7.75%MAPE for

the large two-storey semi. The MAPE of the four archetypes fell within 1% of each other. The

prediction accuracy varied between house types, but not substantially, the range of results can

be seen below.

Table 26 Prediction scores for the elastic net model on the combined dataset.

	Back Transformed		
	R^2	RMSE	MAPE
Full Test Set	0.941 ± 0.005	6.33 ± 0.18	7.28 ± 0.22
Three-Storey Semi	0.946 ± 0.011	6.01 ± 0.51	7.13 ± 0.56
Large two-storey semi	0.934 ± 0.01	5.66 ± 0.65	7.75 ± 0.49
Small Two-Storey Semi	0.93 ± 0.017	6.16 ± 0.43	6.87 ± 0.47
Detached	0.94 ± 0.009	7.28 ± 0.53	7.33 ± 0.36

The results above confirm that an elastic net model can be used to effectively select features from a large group of highly collinear and produce a model that can predict energy use for unseen training examples from multiple different archetypes.

Based on Figure 65 elastic net regression can be used to produce a model lower test error than the stepwise model. This improved accuracy will come at the expense of including additional features compared to the stepwise model. The elastic net regression model produced a model with the same number of features on average as the stepwise model for a model with the same accuracy. Reducing the number of features may not always be a priority for a designer. If elastic net regression is used to select between a large number of collinear features, it is essential to assess the stability of estimated coefficients to check the variance if interpretability is considered important.

4.1.10 Comparing Stepwise and Elastic net models

The mean standardized regression coefficients were plotted with error bars representing standard deviation to assess the stability of the coefficients generated from various models. The coefficients estimated by the stepwise model are shown in Figure 68.
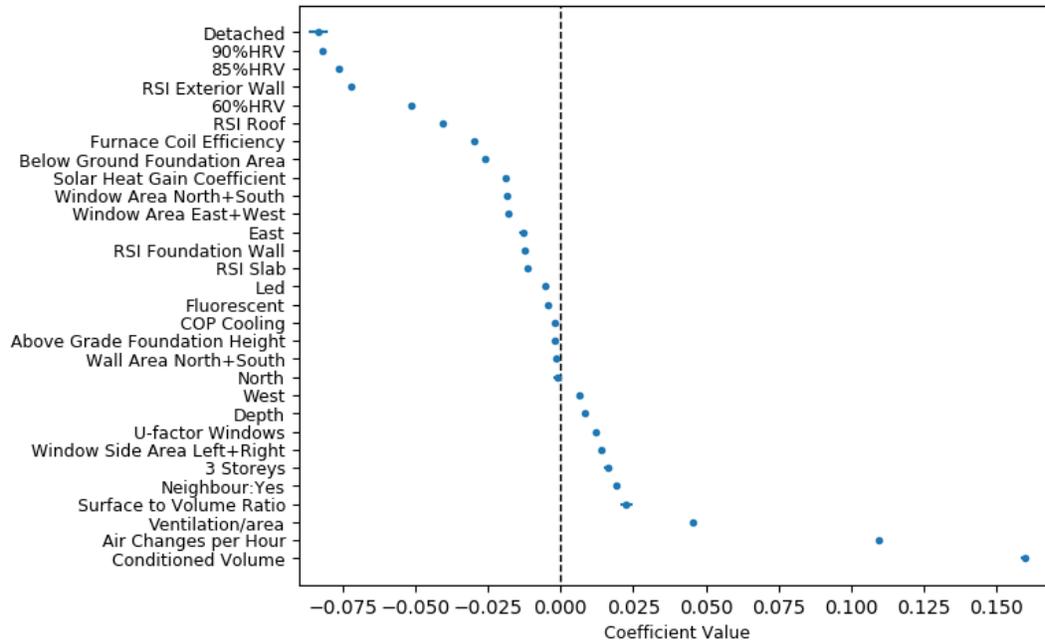
Figure 68 Standardized Regression Coefficients for the stepwise model

The error bars are not visible for most features. the Detached House option shows the highest variance. There appears to be collinearity between Surface to Volume Ratio and the Detached option. The differences in energy use between semi and detached houses are likely related to surface to volume ratio so both features describe similar effects. VIF was not calculated for categorical features during stepwise selection; house type would likely have been removed if this had been considered. A visual inspection of Figure 68 by looking at the magnitude of the error bars in relation to the magnitude of the coefficients suggests that there is no problematic collinearity. The order of the coefficients by magnitude appears to be conserved no matter

where the coefficient estimates fall within the range of expected values shown by the error bars.

Figure 69 shows an elastic net model with 50 features included, corresponding to a tuning parameter(l) of 0.000001.
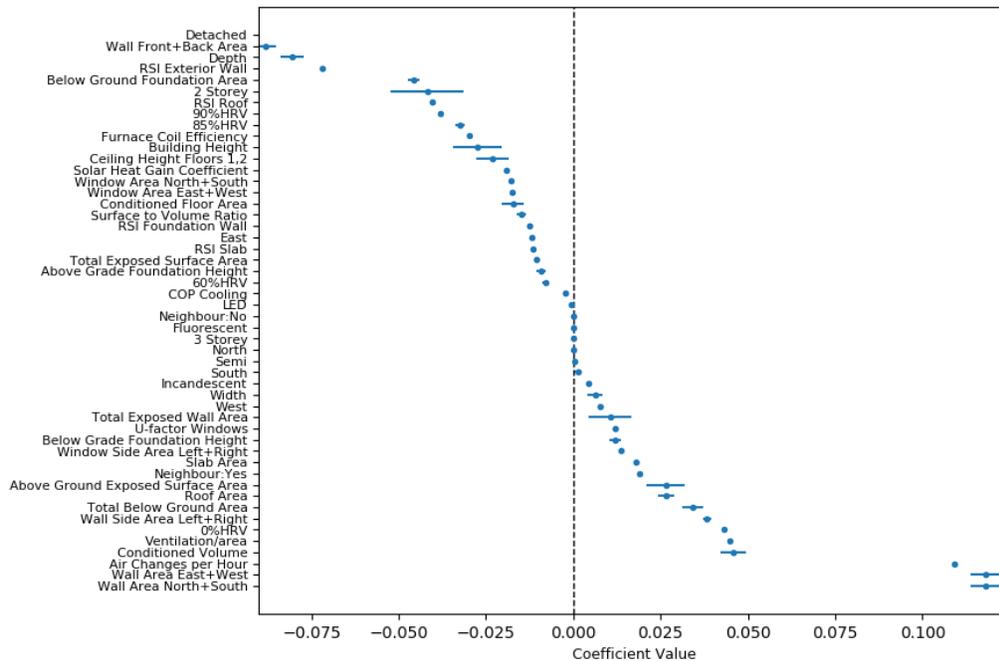
Figure 69 Standardized regression coefficients generated using elastic net, showing high coefficient variance.

This is a model with almost no shrinkage of the coefficients, no coefficients have been reduced to zero. Depending on where the coefficient estimates from each cross-validation split fall within the range of the error bars shown, the order of the absolute value of the coefficients will vary substantially. Generalized geometry features have very high variance, the two-Storey option has the highest variance, along with building height and ceiling height. The combined effect of floor to ceiling height and overall building height should be enough to predict whether a building is two or three storeys with very high accuracy, so this variance would be expected.

There is a lot of variance in all the geometry features in this model. The above figure shows that collinearity will not necessarily always be addressed by elastic net regression. A sufficiently large l value is required to reduce collinearity to an acceptable level to allow the interpretation of the coefficient values. It is difficult to quantify the amount of variance remaining in the model with regularized regression. Pearson correlation and VIF values can not be calculated easily for features that have been shrunk with elastic net. A benefit of the stepwise selection strategy is that the process was automated, and the amount of collinearity was quantified. With the elastic net approach, visual inspection was used to assess the variance in the coefficients. There is likely a way to quantify the amount of collinearity for each l value, this was outside the scope of this research.

The chosen elastic net model is shown in Figure 70, coefficients with a value of 0 are removed from this plot.

The variance in the coefficient estimates has been substantially reduced from the model with effectively no regularization to a model in which a l value of 0.0099 was chosen. The collinearity in the stepwise and the elastic net model with a l value of 0.099 both show very low variance in the standardized regression coefficients.

The combined models developed with stepwise and with elastic net were both effective at creating an accurate and interpretable model. The updated stepwise approach effectively addressed the limitations found in the existing research and was a fully automated process. The elastic net regression was more simple than the stepwise approach and can potentially produce a more accurate model depending on the number of non zero coefficients included. The limitation of elastic net regression in this approach is that it is not as simple to quantify collinearity, this can likely be addressed in future research.

4.1.11 Artificial Neural Networks Combined Dataset.

A preliminary analysis using artificial neural networks was conducted to determine what level of accuracy can be achieved in a black box style model. The pros and cons of using this approach are compared with regression models. Careful tuning is required to optimize the performance of neural networks. Care was required to ensure that no data leakage occurred. In ANN's overfitting is much more likely than in a regression model, so data leakage can have a large effect on the accuracy of the final model than in a regression model.

Careful selection of features is not as important in neural networks because collinearity does not pose any problems. In order to reduce the scope of this analysis, the input features chosen by forward stepwise selection for the combined model were used.

A single hidden layer was selected instead of testing multiple hidden layers. A single hidden layer was the most common structure for ANN's found in the existing research [12]. Only a single hidden layer is required to capture non-linear relationships in the data. The optimal number of nodes in the hidden layer was determined by finding the minimum validation error scores. Models were trained using 40 to 100 hidden nodes, the validation error reached a minimum value and began increasing between 60-70 nodes. After reaching a minimum validation error the training error continued decreasing due to overfitting, and the validation error began increasing. The results of the hyperparameter tuning are shown in Figure 71.
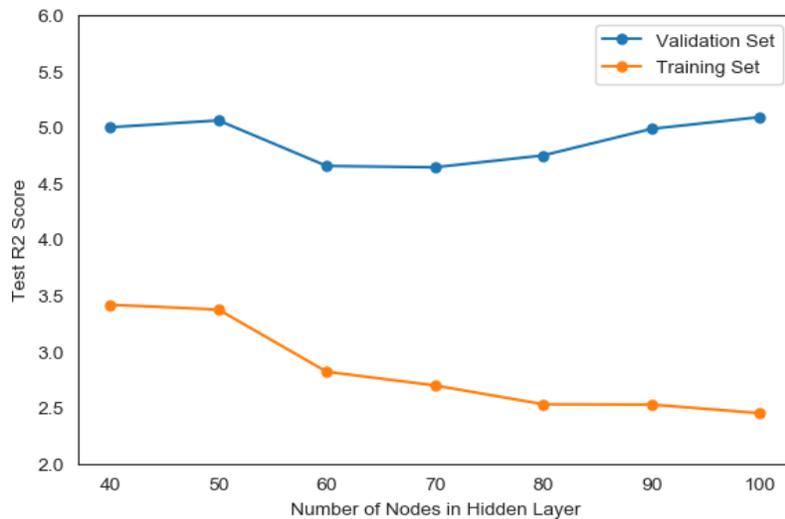
Figure 71 Validation vs. test scores for increasing numbers of hidden nodes.

The number of epochs was determined for each run of the inner cross-validation loop by using an "overfitting call back". The model automatically stopped fitting when the validation error stopped improving. The average number of epochs from these runs was used to train a model on the training + validation set with 70 hidden nodes based on the results found above. This model was trained and tested 10 times, in each instance a different random set of starting weights was used, running the model multiple times gives an estimate of the performance of

134

the neural network. A deterministic model could be developed by fixing the random seed used in generating the initial weights. The mean prediction scores for the ANN model are reported in Table 27.

Table 27 Artificial Neural Network Results

Subset:	R^2	RMSE	MAPE
Full Test Set	0.977 ± 0.002	4.061 ± 0.18	4.54 ± 0.149
Three-storey Semi	0.976 ± 0.003	3.894 ± 0.263	4.54 ± 0.223
Two-storey Semi 2	0.979 ± 0.001	3.137 ± 0.042	4.19 ± 0.073
Two-storey Semi 1	0.969 ± 0.003	4.286 ± 0.188	5.07 ± 0.136
Detached	0.978 ± 0.002	4.699 ± 0.217	4.37 ± 0.253

The results are summarized in Table 27. The model was able to predict with between 4.19 to 5.07% error on each archetype, with an R^2 between 0.97 and 0.98. The training set was able to achieve an R^2 greater than 0.99.

The residuals and predicted vs. actual plots are compared below for the stepwise model (Figure 73), elastic net model (Figure 72) and ANN model (Figure 74).
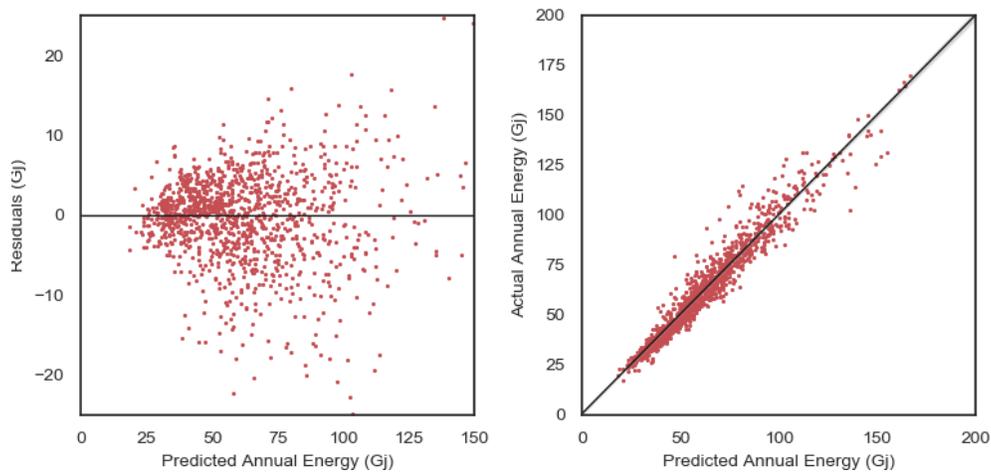
Figure 72 Stepwise model fit

135
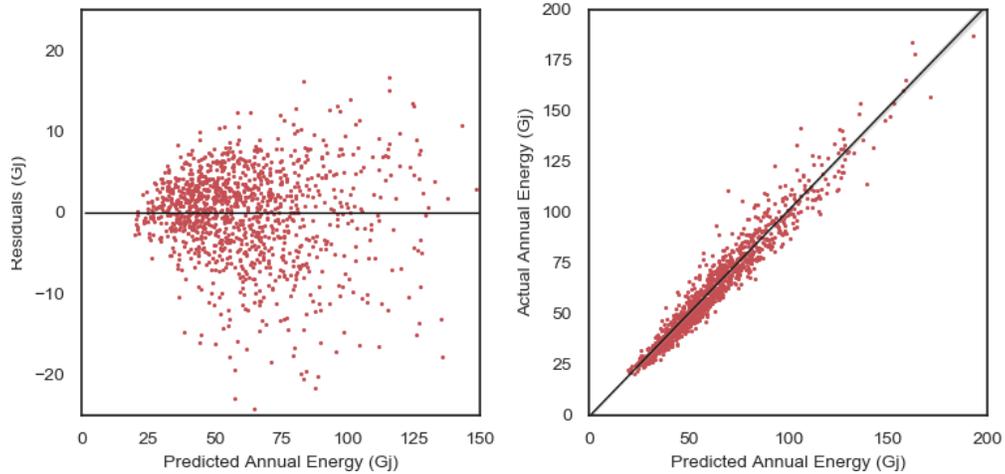
Figure 73 Elastic net model fit
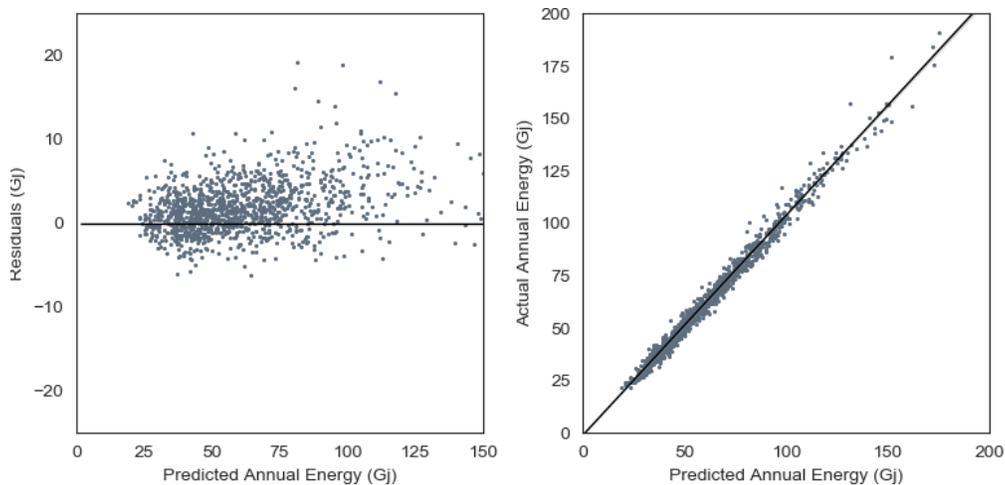
Figure 74 Artificial Neural Network model fit.

The plots for the neural net model show a tighter fit on the predicted vs. actual plots; the fit at higher energy use is significantly better than for regression models. The residual plot also shows smaller residuals and less obvious trends in the ANN model, compared to the regression model. These results suggest that there was non-linear behaviour in the underlying data relationships that could not be fully captured by a linear model, even with the power transformation.

The preliminary analysis using neural networks shows that these models are better suited to fit the relationships found in the synthetic dataset developed for this study. With some tuning of structure and hyperparameters, these models can outperform linear models. The accuracy could further be improved by performing feature selection or adding additional features. The feature set was selected by stepwise regression and was not representative of the best set of features for a more complex model in which collinearity is not a concern.

Although the accuracy was better with the neural net model, the linear regression in this study was only a 2-3% worse. The neural net model can not be interpreted and can not be easily used outside of python. It can not be exported into an excel spreadsheet, for example. The importance of individual features and underlying relationships can not be readily determined. Using elastic net regression and tuning only the l value is a simple methodology that will produce a relatively accurate, sparse and interpretable model.

Building and tuning a neural network requires more expertise than building a regression model. The impact of data leakage, overfitting, and improperly validating the model have more severe consequences with a more flexible algorithm such as a neural network that can perfectly fit the training data. The application of a neural network would be best suited to a problem in which high accuracy is required, and interpretability is not.

5 Further research

Improving the practicality of the model

There is room for future work to be conducted in several areas. Collaboration with industry partners, policymakers, designers and builders should be considered in order to guide the model intent and inform the feature selection process. This will be a key step in developing the practical application of this type of model. More work should be done on extending this model to cover all century houses within Toronto. This research investigated the feasibility of surrogate modelling approaches and identified some effective methodologies. Future research should be able to focus more on the application of surrogate modelling and less on methodology development and data science. Practical application must be considered for future research so that this tool can be implemented. To ensure that this work is practical, validating surrogate models against a database of actual houses would be a useful step in additional research in order to further validate this approach. Incorporating occupant behaviour in models as categorical variables would be a valuable addition to this research.

Researchers should investigate the maximum absolute percent error in addition to mean absolute percent error. The percent error for an individual training example ranged between approximately 37% to 0.008%. 37% error is not acceptable, training examples with high percent error should be investigated to determine why the model is not fitting these houses well. A rule should be developed to determine what is an unrealistic combination of features in an example. This can be used to determine if certain training examples that are outliers which

reduce the accuracy of the model can be removed. Any cleaning up of the dataset should be done very carefully to ensure the model can still predict accurately on all houses that the surrogate model is being used to describe. Removing unrealistic training examples will likely improve the model substantially.

HVAC

An important next step is including HVAC options in the model. This research looked at changing the efficiency of the HVAC system, but different systems should be tested as well. Electric systems such as air source or ground source heat pumps can significantly reduce carbon emissions and energy use compared to natural gas furnaces. Different HVAC systems can be modelled as categorical features. Likely more training examples will be required if multiple HVAC systems are considered.

Occupancy

Incorporating Occupant behaviour should also be considered in future models. Changing the scheduling of setpoints and incorporating natural ventilation, for example, will have an affect on energy use. Setpoint temperatures can be incorporated as a numerical feature, setpoint scheduling would be incorporated as categorical features. Appliance types and use profiles, for example, can be incorporated as categorical features in the model. This would be an improvement from the current archetype modelling approach in which only a single occupancy profile is generally used. A set of occupancy profiles could be developed based on low, medium or high energy use occupancy types. As an example, if a house had a certain set of appliances a certain occupancy profile could be applied.

Neural Networks to Predict Multiple Target Variables

The artificial neural network methodology should be investigated more thoroughly. The artificial neural network model yielded higher accuracy than the multivariate regression model. One of the distinct advantages of using neural networks is that a single neural network can easily be trained on multiple target variables. Instead of total annual energy a neural network can be trained on multiple outputs such as each end-use energy components or monthly energy use. End-use energy components would enable an estimation of carbon emissions by using emissions factors for natural gas and the electrical grid. End-use components would also enable a more detailed analysis of retrofit solutions and their effects.

As an example, upgrading the light type from incandescent to LED would reduce the electrical use but would increase the heating demand. If a natural gas furnace is being used for heating this would mean that switching to a more efficient light type would increase natural gas use to compensate. Depending on the emissions factor for the electricity grid, the result could be a substantial increase in carbon emissions from a slight decrease in total energy use. For this reason it is important to consider carbon and to consider end-use components of energy use.

Training a neural network on monthly energy would enable a more robust calibration of a surrogate model to an existing home. Calibrating an energy model using annual energy use is not considered an adequate measure of how well a houses actual energy performance is being captured.

Elastic Net Feature Selection

A method to quantify collinearity during elastic net regression should be investigated. A quantifiable measure of collinearity this would allow automation of the elastic net regression methodology for feature selection. In this research the only benefit stepwise regression had over elastic net was that the stepwise feature selection was fully automated, and the collinearity was quantified through a VIF calculation.

Optimization

There are many opportunities for future work in optimizing retrofit strategies using full multi-objective optimizations. Current archetype research has been limited by model run times. Surrogate modelling allows full optimizations to be performed in substantially shorter time frames. Using a surrogate model could be used to run optimizations with various energy use targets and baseline conditions.

Modifying Energy Models (*.idf*s)

A script was written to randomly modify geometry for EnergyPlus *.idfs* in order to create four archetypes with a range of sizes and aspect ratios. The basic proportions of each archetype on each axis were maintained and stretched by a factor. The basement heights, ceiling heights and above ground basement height were randomized. More work could be done in this area to develop a larger number of random house geometries with less predictable dimensions. This could be used to create a model that would more comprehensively cover the housing stock.

Field Research

More work can be done in determining baseline archetypes and applying a surrogate modelling methodology to model other house types such as modern, 1970s OBC and wartime. The data collection process should be updated. More data was collected than required during the field study; only data that is required for the modelling process should be recorded to simplify and speed up this process. Only exterior dimensions, window areas and floor heights need to be measured, and only energy modelling inputs should be considered to save time. Envelope related information can not be reliably determined through non-destructive field study in every house. Google earth can also be used to roughly measure a large number of houses using 3d measurement tools, more work could be done to investigate ways to streamline the data collection process, this process could potentially be automated. Future work could also investigate the application of the set of homes described by the CSDDRD [74], could be used to develop new archetype models in Toronto.

User Interface

Creating a user interface for surrogate models would be an important next step to consider. The end-user would have to be considered for a model interface. The application of surrogate modelling is more suited for groups with some knowledge of building systems and performance. This tool could be helpful as a design stage tool for an architect, or a builder who is looking to determine what energy savings are possible for a retrofit project based on the proposed design. If a tool was developed for homeowners, the inputs would have to be carefully explained, this would be more of an educational tool than a design tool.

Canadian Housing Stock Model

There are limits to the scope of a single surrogate model, each model can only represent homes in the specific climate zone or city that was used to run the base energy models.

Once a surrogate model has been developed for a certain subset of the housing stock such as century homes in Toronto.

6 Conclusions

This research provides an important preliminary exploration into the application of surrogate modelling for bottom-up archetype development in Toronto. Models were developed to determine whether it is feasible for multiple archetypes to be described by a single model and whether a decrease in prediction accuracy would be expected. EnergyPlus was used to simulate 6000 houses using performance parameters and geometry feature ranges through field study. Models were developed using simple multivariate linear regression, elastic net regression and artificial neural networks to enable rapid prediction of energy use for four distinct single-family house archetypes in Toronto, ON.

A novel Forward stepwise selection methodology with a built-in collinearity threshold was used to select the best combination of features for two different two-storey semis, a three-storey semi and a semi-detached archetype. The entire surrogate modelling process with the stepwise methodology was automated. This process that was automated included creating the energy models for the dataset, determining the optimal feature transformations, preprocessing, selecting features and validating the model. The model produced for each house type predicted annual energy use with 7.02 to 7.53% error on average. A combined model that contained all four archetypes was developed to determine if a single model can replace multiple models. This model predicted annual energy use with 7.03% error on average. These results confirm that a single surrogate model can describe multiple single-family house archetypes. Models were also trained using three archetypes and tested on a fourth in order to assess the accuracy of a

combined model, testing on archetypes not included in the training set. These models were developed to simulate the accuracy that can be expected when using a surrogate model to predict the energy use for any single-family century home in Toronto. These models predicted energy use with between 7.8 and 14.6% error on average.

Currently, the surrogate modelling methodology is more applicable as a model of a large set of houses, the individual error for a single training example was as high as 37%. The worst prediction accuracy is potentially associated with unrealistic combinations of input features. On the other hand, training examples with more realistic combinations of parameters had percent error as low as .0082%. More work needs to be done to determine why the outlying data points exist and whether they can be removed to obtain a more accurate model.

Using the stepwise feature selection the R^2 improved from 0.84 to 0.94 and the variance inflation factor stayed below 5 for a model with 11 out of 45 potential geometry features included. Elastic net regression was tested as an alternative to the proposed stepwise regression. this methodology was simpler to implement and equally effective at producing a sparse model with high accuracy and low collinearity. More work should be done to quantify collinearity for each l value to allow automation of the surrogate modelling process.

An artificial neural network was also built to determine the accuracy that is possible with a more flexible learning algorithm. This model predicted energy use with a MAPE of 4.5% suggesting that if accuracy is the only objective, an artificial neural network will provide a more suitable model than multivariate regression. A future researcher should consider whether it an

interpretable model is required and whether the compromise in prediction accuracy is worth it given the research objectives.

 A sample size analysis was conducted to check if an adequate number of training examples were used for each stepwise model. The individual archetype models could have had better accuracy if more samples were included in training. The test scores were still improving at 1200 samples. The accuracy of the combined model reached a maximum after only 2000 out of 6000 total samples and this model did not appear to require more total samples than the individual models. Using a single surrogate model for all four archetypes included in this research reduced the number of simulated training examples that were be required to model each archetype by a factor of 3 to 4 compared to the number of samples required to model the archetypes individually. Combining multiple archetypes into one model did not cause a decrease in accuracy and instead was found to substantially decrease the time required to develop models in both dataset and model development. **Using the methodology outlined in this research, accurate and interpretable models can be developed quickly to support the application of surrogate modelling to improve the current archetype modelling approach. This methodology increases the feasibility of using surrogate modelling to accurately model large subsets of the housing stock and enables more complex analysis that was previously not viable.**

Glossary

Feature	An input that describes a specific building performance parameter in the model. For example window to wall ratio, wall RSI, Air changes per hour.
Training Example	An EnergyPlus model that represents a potential combination of feature values. The set of training examples is the synthetic dataset, this is used to build the surrogate model.
Base Feature	A feature that is directly input to EnergyPlus to create the training examples.
Combined Feature	A linear combination of multiple input features.
Target Variable	The value that the learning algorithm is trying to estimate using the features values. This is the output of the learning algorithm.
Validation Set	Data split used to estimate test error to allow comparison between model types and tune hyperparameters. Using this set to choose the best hyperparameters means that it is not longer an independent test set. Validation accuracy is an optimistic estimate of the true error.
Training Set	data split used to generate the model by fitting the relationship between the input and output features.
Testing Set / Holdout Set	Independent Data split used to report the accuracy of the model developed with the training+validation set. This data has not been used to build the model.
Mean Absolute Percent Error (MAPE)	The mean of the percent error for each training example. Also called average error.
Learning Algorithm	A predictive algorithm that estimates an output based on a set of inputs by training it on a set of examples.
Multivariate Regression (MAPE)	Linear equation with multiple inputs. $Y = b1 + b1x1 + b2x2...$
Artificial Neural Network (ANN)	Learning algorithm that uses nodes to emulate the neurons in a brain. Considered a universal approximator that can model any linear or non linear function if the correct structure is used.
Elastic Net	Regression with shrinkage penalty used for feature selection as well
Combined Model	The surrogate model containing all four archetypes (6000 training examples)
Individual Model	Surrogate model for an individual archetypal house (1500 training examples).
Hyperparameter	A parameter that changes how a learning algorithm functions.
Tuning Parameter	An example of a hyperparameter in elastic net regression that determines how much the coefficients are shrunk.
Standardization	Puts features on the same scale by subtracting the mean and dividing by standard deviation.
Cross-validation	Splits dataset into equal components, validation and testing and determines average performance of model. K-fold cross validation is used in this research.

Nested cross-validation	Uses a cross validation loop to evaluate accuracy and an inner loop to tune model hyperparameters.
Fold	A set of training examples within the cross-validation loop.
Split	A specific set of training examples sampled from the full dataset.
Latin Hypercube Sampling	Sampling plan that creates random input values evenly distributed between 0 and 1 used in conjunction with the feature ranges to create a random set of training examples.
Categorical Feature	A feature that represents an option, described by a binary value 0 or 1.
Numerical Feature	A continuous numerical feature that represents a range of values.
Back-transform	Converting log transformed output into units of Gj by raising 10 to the exponent of the value.
Actual Energy	Refers to simulated annual energy use using EnergyPlus.

Appendices

A1: Cross-Validation Results

Elastic Net Nested Cross-Validation Results (Combined Model)

Set		Log Transformed Scores			Back Transformed Scores		
		R2	RMSE	MAPE	R2	RMSE	MAPE
Fold 1	Validation	0.944	0.042	1.862			
	Test	0.951	0.039	1.757	0.946	6.136	7.047
	Three Storey Semi	0.954	0.038	1.713	0.949	6.027	6.840
	Large Two-Storey Semi	0.929	0.042	1.987	0.922	5.717	7.888
	Small Two-Storey Semi	0.954	0.036	1.609	0.942	5.942	6.395
	Detached	0.957	0.039	1.707	0.953	6.863	7.036
Fold 2	Validation	0.946	0.041	1.842			
	Test	0.940	0.043	1.903	0.934	6.510	7.506
	Three Storey Semi	0.936	0.044	1.951	0.932	6.570	7.654
	Large Two-Storey Semi	0.950	0.037	1.837	0.949	4.614	7.209
	Small Two-Storey Semi	0.925	0.045	1.956	0.910	6.551	7.611
	Detached	0.938	0.045	1.856	0.933	7.723	7.506
Fold 3	Validation	0.946	0.041	1.840			
	Test	0.945	0.042	1.877	0.939	6.631	7.535
	Three Storey Semi	0.942	0.042	1.863	0.944	6.174	7.354
	Large Two-Storey Semi	0.936	0.042	2.033	0.923	6.041	8.187
	Small Two-Storey Semi	0.946	0.039	1.764	0.943	5.980	6.971
	Test_detached	0.942	0.044	1.852	0.932	8.108	7.650
Fold 4	Validation	0.945	0.041	1.840			
	Test	0.948	0.041	1.861	0.942	6.310	7.392
	Three Storey Semi	0.943	0.042	1.874	0.935	6.557	7.374
	Large Two-Storey Semi	0.942	0.044	2.159	0.936	6.152	8.507
	Small Two-Storey Semi	0.951	0.038	1.714	0.943	5.636	6.706
	Detached	0.947	0.040	1.681	0.945	6.774	6.895
Fold 5	Validation	0.945	0.041	1.842			
	Test	0.945	0.041	1.804	0.938	6.330	7.201
	Test_semi3	0.958	0.036	1.590	0.959	5.385	6.323
	Test_semi2	0.942	0.042	1.978	0.932	6.167	7.834
	Test_semi1	0.935	0.041	1.741	0.918	6.508	6.873
	Test_detached	0.934	0.044	1.896	0.932	7.107	7.730
Mean	Validation	**0.945**	**0.041**	**1.845**			
	Test	**0.946**	**0.041**	**1.840**	**0.940**	**6.383**	**7.336**

	Three Storey Semi	0.947	0.040	1.798	0.944	6.143	7.109
	Large Two-Storey Semi	0.940	0.042	1.999	0.932	5.738	7.925
	Small Two-Storey Semi	0.943	0.040	1.757	0.931	6.123	6.911
	Detached	0.944	0.042	1.798	0.939	7.315	7.364
	Validation	0.001	0.000	0.009			
	Test	0.004	0.002	0.059	0.005	0.191	0.208
SD	Three Storey Semi	0.009	0.003	0.145	0.011	0.486	0.529
	Large Two-Storey Semi	0.008	0.003	0.116	0.011	0.654	0.482
	Small Two-Storey Semi	0.012	0.003	0.126	0.016	0.394	0.448
	Detached	0.009	0.003	0.097	0.010	0.578	0.375

Stepwise Feature Selection Nested Cross-Validation Results (Combined Model)

		Log Transformed			Back Transformed		
		R2	RMSE	MAPE	R2	RMSE	MAPE
Fold 1	Test	0.954	0.038	1.683	0.948	6.067	6.756
	Three Storey Semi	0.957	0.036	1.660	0.951	5.885	6.685
	Large Two-Storey Semi	0.934	0.040	1.873	0.926	5.567	7.368
	Small Two-Storey Semi	0.956	0.035	1.553	0.942	5.919	6.193
	Detached	0.958	0.038	1.635	0.953	6.895	6.748
Fold 2	Test	0.944	0.042	1.827	0.935	6.439	7.218
	Three Storey Semi	0.942	0.042	1.882	0.939	6.198	7.410
	Large Two-Storey Semi	0.958	0.034	1.654	0.953	4.425	6.457
	Small Two-Storey Semi	0.931	0.043	1.900	0.911	6.532	7.466
	Detached	0.937	0.045	1.845	0.929	7.919	7.435
Fold 3	Test	0.949	0.040	1.786	0.942	6.468	7.179
	Three Storey Semi	0.948	0.040	1.775	0.952	5.726	7.046
	Large Two-Storey Semi	0.945	0.039	1.833	0.928	5.842	7.342
	Small Two-Storey Semi	0.953	0.037	1.671	0.947	5.741	6.654
	Detached	0.940	0.044	1.870	0.930	8.228	7.691
Fold 4	Test	0.951	0.040	1.794	0.942	6.306	7.140
	Three Storey Semi	0.947	0.040	1.835	0.939	6.309	7.286
	Large Two-Storey Semi	0.946	0.043	2.013	0.936	6.153	7.891
	Small Two-Storey Semi	0.954	0.037	1.650	0.944	5.608	6.482
	Detached	0.948	0.040	1.665	0.942	6.990	6.832
Fold 5	Test	0.948	0.040	1.718	0.940	6.228	6.867
	Three Storey Semi	0.963	0.034	1.523	0.964	5.026	6.135
	Large Two-Storey Semi	0.949	0.040	1.793	0.935	6.016	7.064
	Small Two-Storey Semi	0.939	0.039	1.679	0.924	6.265	6.653
	Detached	0.933	0.045	1.869	0.927	7.351	7.587
Mean	Test	**0.949**	**0.040**	**1.762**	**0.942**	**6.302**	**7.032**
	Three Storey Semi	**0.951**	**0.038**	**1.735**	**0.949**	**5.829**	**6.912**
	Large Two-Storey Semi	**0.946**	**0.039**	**1.833**	**0.936**	**5.600**	**7.224**
	Small Two-Storey Semi	**0.947**	**0.038**	**1.691**	**0.933**	**6.013**	**6.689**
	Detached	**0.943**	**0.043**	**1.777**	**0.936**	**7.477**	**7.259**
SD	Test	**0.003**	**0.001**	**0.059**	**0.004**	**0.163**	**0.207**
	Three Storey Semi	**0.008**	**0.003**	**0.144**	**0.010**	**0.506**	**0.515**
	Large Two-Storey Semi	**0.008**	**0.003**	**0.130**	**0.011**	**0.692**	**0.523**
	Small Two-Storey Semi	**0.011**	**0.003**	**0.127**	**0.016**	**0.381**	**0.473**
	Detached	**0.010**	**0.003**	**0.117**	**0.011**	**0.581**	**0.438**

Small Two-Storey Semi Data collection

Geometry

House:	House 1	House 2	House 3	House 4
Floor to Ceiling Basement	2.12	2.02	2.11	2.19
Floor to Ceiling Main Floor	2.50	2.54	2.54	2.38
Floor to Ceiling Second Floor	2.46	2.43	2.54	2.41
Floor Thickness	0.20	0.27	0.25	0.30
Above Grade Basement Height	1.10	1.15	1.12	1.08
Ground Floor Width	4.34	4.17	4.89	4.90
Second Floor Jog Width	3.59	3.52	4.20	4.19
Depth Main Floor	11.78	11.61	11.57	11.40
Depth Second Floor	12.92	12.77	12.99	12.82
Porch Roof Depth	2.32	2.23	2.26	2.29
Porch Roof Height	3.55	3.80	3.70	3.59
Building Height (Flat Roof)	6.29	6.57	6.40	6.04
Building Height (To Peak Roof)	7.49	7.82	7.53	7.50
Attic Roof Depth	6.43	5.57	5.25	5.20
Flat Roof Length	6.49	7.20	7.74	7.62
Rear Addition Height1	na	na	3.68	na
Rear Addition Height2	na	na	3.94	na
Rear Addition Depth	na	na	3.40	na
Rear Addition Width	na	na	2.40	na

Window areas

	house 1	house 2	house 3	house 4	average
Basement Side	0.356128	0.304	0.686596	0.84494	0.547916
Wall area basement side	12.958	13.3515	12.9584	12.255	12.88073
WWR Side	0.027483	0.022769	0.052985	0.068947	0.042538
Basement Back	0.301935	0.304	0.34	0.42247	0.342101
Wall area basement back	4.774	4.79205	5.4768	5.2675	5.077588
WWR basement back	0.063246	0.063438	0.06208	0.080203	0.067375
Main Front	1.300643	1.8539	1.661932	1.3312	1.536919
Wall area Main Front	11.9784	11.6551	13.692	12.9458	12.56782
WWR Main Front	0.108582	0.159063	0.12138	0.102829	0.12229
Main Side		1.3688	1.858061	1.3312	1.519354
wall area side main	32.5128	32.47317	32.396	30.1188	31.87519
WWR Side Main		0.044007	0.060844	0.046242	0.047666
Main Back	0.838708	0.726	0.37823	0.4836	0.606635
wall area back main	11.9784	11.6551	13.692	12.9458	12.56782
WWR Back	0.070018	0.06229	0.027624	0.037356	0.048269
Second Fron	1.546449	2.3474	3.178058	1.8432	2.228777
wall area front second	9.760338	9.47936	11.7656	11.16102	10.54158

	C1	C2	C3	C4	C5
WWR Front second	0.158442	0.247633	0.270114	0.165146	0.211427
Second Side	1.110966	1.012	0.477418	1.94754	1.136981
Wall area side second	35.11384	34.40038	36.3748	34.16797	35.01425
WWR Side second	0.031639	0.029418	0.013125	0.056999	0.032472
Second Back	0.697418	0.6545		0.9344	0.762106
Wall area back second	11.79612	11.22173	13.692	13.0585	12.44209
WWR Back second	0.059123	0.058324		0.071555	0.061252
avg side	0.018206	0.033466	0.036977	0.053875	0.040169
avg front	0.107386	0.162046	0.15646	0.108067	0.133597
avg back	0.064384	0.060881	0.021857	0.058854	0.056862

Structure

Sides and back main	**Sructure**	Double Wythe	Wood Frame	double wythe	Wood frame
	Cladding		Brick Face		Face brick
	Insulation	3/4" foam		rsi 1.3	
Front main	**Sructure**	double wythe	Wood frame	double wythe	Wood frame
	Cladding		Brick face		Face brick
	Insulation	3/4"xps		rsi 1.3	
Sides and back second	**Sructure**	wood frame	Wood frame	double wythe	Wood frame
	Cladding	aluminum	Cedar shakes		Aluminum siding
	Insulation	blown in 3.5" old	fib batt 3.5"		1"xps
Front second	**Sructure**	wood frame	Wood frame	double wythe	Wood frame
	Cladding	aluminum	Cedar shakes		Aluminum siding
	Insulation	blown in 3.5" old	fib batt 3.5		
Roof insulation	**Flat**			rsi 3.24(18.4) effective 3.01(17)	
	Attic			rsi 3.24(18.4) effective 3.01(17)	
Basement Slab	Structure	Concrete	Concrete	Concrete	Concrete
	Insulation				
Basement foundation	**Sructure**	triple wythe	triple wythe	triple wythe	triple wythe
	Insulation	none			
Exposed floor	Insulation				

Large Two-Storey Semi Data collection

Geometry

House:	House 1	House 2	House 3	House 4	House 5
Floor to Ceiling Basement	1.97	2.28	2.42	1.93	1.80
Floor to Ceiling Main Floor	2.60	2.75	2.62	2.72	2.40
Floor to Ceiling Second Floor	2.45	2.48	2.46	2.60	2.56
Floor Thickness	0.21	0.24	0.26	0.24	0.24
Above Grade Basement Height	1.20	1.10	1.14	1.14	1.36
Ground Floor Width	4.33	4.50	4.57	4.85	5.09
Second Floor Width	4.33	4.50	4.57	4.85	5.09
Depth Main Floor	12.28	15.76	13.39	12.30	12.87
Depth Second Floor	12.28	15.76	13.39	12.30	12.87
Porch Roof Depth	2.22	1.98	2.91	2.47	2.60
Porch Roof Height	2.58	2.52	3.70	3.90	3.84
Building Height (Flat Roof)	6.68	6.66	6.61	6.77	6.82
Building Height (To Peak Roof)	8.80	8.95	9.03	9.16	8.57
Attic Roof Depth	4.30	4.20	4.60	5.96	3.70
Flat Roof Length	8.43	12.01	7.82	6.34	9.43
Rear Addition Height1	3.97	6.66	na	na	2.28
Rear Addition Height2	4.43	6.66	na	na	4.01
Rear Addition Depth	2.75	2.90	na	na	4.17
Rear Addition Width		2.41	na	na	2.76

Window Areas

	house 1	house 2	house 3	house 4	house 5	Average
Basement Side	0.438709	0.614192	0.469676	0.659999	1.028385	0.642192
Wall area basement side	14.736	17.336	15.2646	14.022	17.5032	15.77236
WWR Side	0.029771	0.035429	0.030769	0.047069	0.058754	0.040716
Basement Back	0	0.731611	0.234838	0.278709	0.318064	0.312645
Wall area basement back	5.196	4.95	5.2098	5.529	6.92648	5.562256
WWR basement back		0.1478	0.045076	0.050409	0.04592	0.056208
Main Front	1.498062	1.561287	1.401933	0.516128	2.341157	1.463713
Wall area Main Front	12.2972	13.455	13.08391	14.3463	13.45571	13.32762
WWR Main Front	0.121821	0.116038	0.107149	0.035976	0.17399	0.109826
Main Side	4.042895	1.755727	1.025804	1.4329	2.750317	2.201529
wall area side main	34.8752	47.1224	38.33557	36.3834	34.00254	38.14382
WWR Side Main	0.115925	0.037259	0.026759	0.039383	0.080886	0.057717
Main Back	3.484697	3.497412	1.276909	1.576741	1.120643	2.191281
wall area back main	12.2972	13.455	13.08391	14.3463	13.45571	13.32762
WWR Back	0.283373	0.259934	0.097594	0.109906	0.083284	0.164416
Second Fron	1.709674	2.2192	1.678061	1.6129	2.308641	1.905695
wall area front second	12.2972	13.455	13.08391	14.3463	13.45571	13.32762
WWR Front second	0.13903	0.164935	0.128254	0.112426	0.171573	0.142988
Second Side	1.300965	2.200044	1.424513	0.838708	1.987093	1.550265

Wall area side second	33.0332	42.85144	36.19317	34.932	35.99739	36.60144
WWR Side second	0.039384	0.051341	0.039359	0.02401	0.055201	0.042355
Second Back	0.922579	1.944888	1.00645	1.233546	1.237417	1.268976
Wall area back second	11.6477	12.2355	12.35271	13.774	14.24512	12.85101
WWR Back second	0.079207	0.158955	0.081476	0.089556	0.086866	0.098745
avg side	0.069969	0.042587	0.032519	0.034353	0.065892	0.048543
avg front	0.107677	0.118659	0.098159	0.062213	0.137414	0.126407
avg back	0.15124	0.201495	0.082169	0.0918	0.077284	0.118866

Structure

Sides and back main	**Sructure**	Wood Frame	Wood Frame	Wood Frame	Double Wythe	Double Wythe
	Cladding	Brick Face	Brick Face	Brick Face		Angel Stone
	Insulation					
Front main	**Sructure**	Wood frame	Wood frame	Wood frame	Double Wythe	Double Wythe
	Cladding	Brick face	Brick face	Brick face		Angel Stone
	Insulation					na
Sides and back second	**Sructure**	Wood Frame	Wood Frame	Wood Frame	Double Wythe	Double Wythe
	Cladding	Brick Face	Brick Face	Brick Face	na	Angel Stone
	Insulation					
Front second	**Sructure**	Wood frame	Wood frame	Wood frame	Double Wythe	Double Wythe
	Cladding	Brick face	Brick face	Brick face		Angel Stone
	Insulation					
Roof insulation	**Flat**		3"batt		5"	5"
	Attic		3" batt		r60 blown fib glass	
Basement Slab	Structure	Concrete	Concrete	Concrete	Concrete	Concrete
	Insulation					
Basement foundation	**Sructure**	triple wythe	triple wythe	triple wythe	triple wythe	triple wythe
	Insulation		r12			fib glass 3.5

Survey Results

Barriers to performing retrofits ranked on scale of 1-5 from not an issue to major issue.

House:	a	b	c	d	a	b	c	d	a	b	c	d	a	b	c	d	a	average
Cost	4	1	2	5	4	2.5	1	4	5	2	1	5	1	5	5	5	1	3.15
Time	2	3	5	1	1	1	3	3	2	1	5	4	3	1	3	2	2	2.47
Displacement	2	2	2	3	5	1	2	2	1	4	4	2	4	1	3	4	3	2.65
Disruption	1	4	2	3	3	1	4	1	3	2	3	4	2	1	3	3	4	2.59
Aesthetics	3	5	3	1	3	3	6	5	1	2	2	2	5	1	1	2	5	2.94
Unclear Savings	2	5	1	3	4	3	5	2	4	2	1	2	5	1	3	1	3	2.76

Percent energy reduction goals, budget and building features that should be maintained.

Percent Energy Reduction Goal (%)	Budget goal ($)	Aesthetic features to Maintain?
Unspecified	Unspecified	nothing
50	40,000	Brick facade
25	40,000	Interior trim
75	62,500	brick facade
50	62,500	nothing
75	15,000	front brick
Unspecified	15,000	nothing
Unspecified	10,000	nothing
20	15,000	brick facade, interior detailing
15	10,000	nothing
75	62,500	nothing
Unspecified	Unspecified	brick facade
50	62,500	oak front door
50	10,000	nothing
50	15,000	nothing
25	87,500	brick front
40	15,000	nothing
50	15,000	nothing

Python/eppy Geometry Modification Script

```
#bh=new below grade height (m)
#width_x: new width (m)
#base_z: new basement height (m)
#height_z: new floor to ceiling height, main and second floor (m)
#depth_y: new depth (m)
#fh_z: new basement above grade height (m)


#bh1=original below grade height (m)
#width_x0: original width (m)
#base_z0: original basement height (m)
#height_z0: original floor to ceiling height, main and second floor (m)
#depth_y0: original depth (m)
#fh_z0: original basement above grade height (m)


#geometry modification function
def updategeo():
    bh=base_z-fh_z
    bh1=base_z0-fh_z0
    #update vertices in each zone
    for s in surfaces:

        if s.Zone_Name==basementzone and s.tilt==90 and s.Vertex_2_Zcoordinate==0 or s.Zone_Name==basementzone and s.tilt==180:
            for field in s.fieldnames:
                if 'XCOORDINATE' in field.upper() and type(s[field])==float:
                    s[field]=s[field]*(width_x/width_x0)
                if 'YCOORDINATE' in field.upper() and type(s[field])==float:
                    s[field]=s[field]*(depth_y/depth_y0)
                if 'ZCOORDINATE' in field.upper() and type(s[field])==float:
                    s[field]=s[field]*((bh)/(bh1))

        if s.Zone_Name==basementzone and s.tilt==90 and s.Vertex_2_Zcoordinate!=0 or s.Zone_Name==basementzone and s.tilt==0:
            for field in s.fieldnames:
                if 'XCOORDINATE' in field.upper() and type(s[field])==float:
                    s[field]=s[field]*(width_x/width_x0)
                if 'YCOORDINATE' in field.upper() and type(s[field])==float:
                    s[field]=s[field]*(depth_y/depth_y0)
                if 'ZCOORDINATE' in field.upper() and type(s[field])==float:
                    s[field]=(s[field]-bh1)*(fh_z/fh_z0)+bh

        if s.Zone_Name != basementzone:
            for field in s.fieldnames:
                if 'XCOORDINATE' in field.upper() and type(s[field])==float:
                    s[field]=s[field]*(width_x/width_x0)
                if 'YCOORDINATE' in field.upper() and type(s[field])==float:
                    s[field]=s[field]*(depth_y/depth_y0)
                if 'ZCOORDINATE' in field.upper() and type(s[field])==float:
                    s[field]=s[field]*(height_z/height_z0)
    #update zone origins
    for zone in zones:
        if zone.Name==basementzone:
            zone.Z_Origin *=(bh)/(bh1)
        if zone.Name==firstfloorzone:
            zone.Z_Origin =fh_z
        if zone.Name==secondfloorzone:
            zone.Z_Origin =fh_z+height_z
        if zone.Name==thirdfloorzone:
            zone.Z_Origin =fh_z+2*height_z
            zone.Y_Origin*=depth_y/depth_y0
```

Information Sheet for Residents

Hello!

We are 2 master's students from the Graduate Program in **Building Science at Ryerson University** working in collaboration with **The Pocket Community Association**. We are conducting research on renovations to increase energy efficiency for century homes.

We are creating a series of computer energy models of homes in the Pocket neighbourhood to understand their current energy performance, and determine the effectiveness of various renovation strategies to lower heating and cooling energy usage.

We have identified your home as a candidate for this research and are presenting you with an exciting opportunity to participate!

Many homes built at the turn of the century follow similar blueprints, meaning there are a few common repeating house types (archetypes). Instead of creating an energy model for individual homes, our goal is to create a general model of a several housing archetypes that can accurately describe a significant portion of the pocket neighbourhood.

In order to create the energy models, we need to visit specific homes in the neighbourhood that fit the archetype. If you are interested in participating, we require a short visit (1-2 hours) to your home to collect data. We have a flexible schedule so we should be able to work around your availability.

What data will we collect if you decide to participate?

- Measure the interior and exterior dimensions of your home
- Record window measurements and types
- Sketch the floor layout
- Record exterior and interior wall types
- Take a look at the HVAC system

What is required of you if you decide to participate?

- Be present during our data collection
- Fill out a small questionnaire on your energy use habits
- Allow us access to your hydro and gas bills (ideally for the last 2 years, however, we will accept whatever you have ☺)

The research findings will be shared when completed. This is an opportunity to learn about your energy usage. This study will identify the major sources of energy loss in your home and some renovation strategies to save on your energy bill and help the planet.

If you are interested in participating please let us know, and if you have any questions feel free to ask. Thank you and we hope to hear from you!

Cecilia Skarupa – Cskarupa@ryerson.ca
Cameron Lawrence –Cameron.lawrence@ryerson.ca
Supervisor: Professor Russell Richman, Ph.D, P.Eng. –Richman@ryerson.ca

*Note: All data collected from your home will remain confidential. The data collected will only be used to calculate averages or to determine general construction practices. No information or photos will be included in the research. If you chose to participate, you are permitted to discontinue participation at any time or decline to answer any questions. This study, the questionnaire survey, and the data collection methods have been approved by the Ryerson University Research Ethics Board.

Consent Form

<div align="center">

Ryerson University
Consent Agreement

</div>

You are being invited to participate in a research study. Please read this consent form so that you understand what your participation will involve. Before you consent to participate, please ask any questions to be sure you understand what your participation will involve.

TORONTO HOUSING ARCHETYPE STUDY

INVESTIGATORS: This research is being conducted by Cameron Lawrence and Cecilia Skarupa for the partial completion of their graduate studies in the Building Science program at Ryerson University. The research is being supervised by Dr. Russell Richman from the Department of Architectural Science at Ryerson University.

This study is not funded.

If you have any questions or concerns about the research, please feel free to contact:
Cameron Lawrence – cameron.lawrence@ryerson.ca or 416 879 5759
Russell Richman - richman@ryerson.ca or 416-979-5000 ext.
Cecilia Skarupa -Cecilia.skarupa@ryerson.ca

PURPOSE OF THE STUDY:
The purpose of this study is to categorize the house types in the pocket neighbourhood, and determine the effectiveness of various retrofit strategies to achieve high levels of energy efficiency for these homes. 4 different house types(archetypes) will be investigated, data will be collected in order to create an energy model of each house type. A model of the existing house performance will be generated and compared to further models in which various retrofit strategies are tested to increase energy efficiency. This study will focus on century detached and semi-detached homes. The study requires measurements inside 4 homes for each archetype considered to collect sufficient data. 20 people with homes matching the archetype descriptions will be recruited for this study.

WHAT YOU WILL BE ASKED TO DO:
If you volunteer to participate in this study, you will be asked to do the following things:

Data will be collected from your home including housing dimensions, interior wall locations, and heating system information. The data will be used to prepare computer energy models to test energy efficiency and the effectiveness of various renovation strategies.
You will be asked to:

- Fill out a questionnaire survey about your home, the heating system, and appliances (for example: In what room is the thermostat located? How is your home heated? What appliances are typically used? At what time of the day are they used?)
- Provide access to the outside and inside of your home so that the researcher can take measurements and photographs and document the locations of interior walls
- Provide access to the heating system of the home and any manuals for the heating system (i.e. furnace manuals)
- Provide access to utility bills for a two year timespan, these can be obtained directly through the utility provider. Participant can choose to remove identifying/personal information from this data. Energy data will be copied into a spread sheet stored in the Ryerson google drive, utility bills will be electronically deleted once the energy data has been collected.

The research will take place at your home for an expected duration of two (2) hours. The study will be scheduled at a time when you will be home and can be present for the duration of the study.

POTENTIAL BENEFITS:

Participants that participate in this study will be able to better understand the energy profile of their home and how their usage relates to the community average. The participants will be provided with the most effective renovation designs to reduce energy use and improve comfort in their homes. This will simplify the renovation process for the homeowner if they are interested, key areas for improvement that will offer the greatest reduction in energy use will be clearly identified.

A potential benefit to society is the potential to use this research to achieve reductions in energy use and greenhouse gas emissions if the proposed renovations can be applied to multiple homes.

The potential benefit to science is a framework that can be used to conduct similar research in other neighbourhoods in Toronto and abroad.

I cannot guarantee, however, that you will receive any benefits from participating in this study.

WHAT ARE THE POTENTIAL RISKS TO YOU AS A PARTICIPANT:
You may experience discomfort with the researcher entering your home or property and collecting data and taking photographs. You may also experience discomfort in answering the survey questions. To minimize the risk of discomfort, the researcher will not make public any information specific to your home and your identity and address will remain confidential. The data collected will be averaged or used to make generalizations about construction practices. Any photographs taken will be used for the researcher's reference and will not be made public. If you begin to feel uncomfortable you may discontinue participation in the study. If you are uncomfortable with photographs being taken you may inform the researcher and photographs will not be taken.

CONFIDENTIALITY:
You will not be asked to identify yourself for the purpose of the study. All information gathered will be confidential. Your address and information specific to your home will not be reported in the study. The data collected from your home will only be used to determine averages or to make generalizations about construction practices. All collected data will remain anonymous in the report and be made available only to researchers and supervisor.

DATA STORAGE
All data will be stored on the Ryerson Google Drive. Access will be provided only to the researchers and faculty supervisors. No identifying data will be stored past the conclusion of the project. Data will be electronically erased.

DATA DISSEMINATION
Data will be used to generate scholarly research articles and two dissertations. All dissemination is public and participants will be directed to all publications generated.

VOLUNTARY PARTICIPATION AND WITHDRAWAL:
Participation in this study is completely voluntary. You can choose whether to be in this study or not. If any question makes you uncomfortable, you can skip that question. If you choose to stop participating, you may also choose to not have your data included in the study. Your choice of whether to participate or not will not influence your future relations with Ryerson University or the investigators (Cameron Lawrence and Cecilia Skarupa) involved in the research.

QUESTIONS ABOUT THE STUDY: If you have any questions about the research now, please ask. If you have questions later about the research, you may contact.

Cameron Lawrence – cameron.lawrence@ryerson.ca or 416 879-5759
Russell Richman - richman@ryerson.ca or 416-979-5000 ext.
Cecilia Skarupa -Cecilia.skarupa@ryerson.ca 647 923-5413

This study has been reviewed by the Ryerson University Research Ethics Board. If you have questions regarding your rights as a participant in this study please contact:

> Research Ethics Board
> c/o Office of the Vice President, Research and Innovation
> Ryerson University
> 350 Victoria Street
> Toronto, ON M5B 2K3
> 416-979-5042
> rebchair@ryerson.ca

TORONTO HOUSING ARCHETYPE PROJECT

CONFIRMATION OF AGREEMENT:

Your signature below indicates that you have read the information in this agreement and have had a chance to ask any questions you have about the study. Your signature also indicates that you agree to participate in the study and have been told that you can change your mind and withdraw your consent to participate at any time. You have been given a copy of this agreement. You have been told that by signing this consent agreement you are not giving up any of your legal rights.

_____
Name of Participant (please print)

_____   _____
Signature of Participant   Date

_____   _____
Signature of Investigator   Date

PHOTOGRAPHY CONSENT AGREEMENT:

Your signature below indicates that you have read the information above and give researchers consent to take photos in the participants home. The photos will include, HVAC system, and/or building envelope. They are for the investigators and supervisors reference only, and will not be included in any paper or presentation, so the participant confidentiality will be maintained. The photos will be stored on Ryerson Google Drive until two years after the end of the project, at which point they will be deleted.

_____
Name of Participant (please print)

_____   _____
Signature of Participant   Date

_____   _____
Signature of Investigator   Date

RYERSON UNIVERSITY
Questionnaire Survey

TORONTO HOUSING ARCHETYPE PROJECT

1) Have you made any renovations to your home? (Please circle)   YES   NO
   If yes, please list the renovations made and when they occurred:
   _____
   _____
   _____
   _____

2) Are you aware of any renovations made by previous owners?   YES   NO
   If yes, please list the renovations made:
   _____
   _____
   _____
   _____

3) How is your home heated?
   _____
   _____

4) Where is the thermostat located?
   _____
   _____

5) What temperature set point do you maintain for heating and cooling, fill in an estimate for temperature set points on your thermostat at each time interval during the heating and cooling season. Include below the estimated date you typically start using your heating and cooling system.

Heating							
Weekday hours	6am-9am	9am-1pm	12-5pm	5pm-9pm	9pm-6am	never	Rarely
Set point: (˚C)							
Weekend hours	6am-9am	9am-1pm	12-5pm	5pm-9pm	9pm-6am	never	Rarely
Set point: (˚C)							
Cooling							
Weekdays hours	6am-9am	9am-1pm	12-5pm	5pm-9pm	9pm-6am	never	Rarely
Set point: (˚C)							
Weekend hours	6am-9am	9am-1pm	12-5pm	5pm-9pm	9pm-6am	never	Rarely
Set point: (˚C)							

	Start date
Heating	
Cooling	

6) When are major appliances typically used in your house? (fill in the table below by including the duration of use for each time interval, leave blank or 0 if the appliance is not used at this time. Circle the Rarely/Never option if this applies instead)

Kitchen Fan						
Weekdays	6am-9am	9am-1pm	12-5pm	5pm-9pm	9pm-6am	
Hours/time interval						Rarely/Never
Weekend	6am-9am	9am-1pm	12-5pm	5pm-9pm	9pm-6am	
Hours/time interval						Rarely/Never
Kitchen Stove						
Weekdays	6am-9am	9am-1pm	12-5pm	5pm-9pm	9pm-6am	
Hours/time interval						Rarely/Never
Weekend	6am-9am	9am-1pm	12-5pm	5pm-9pm	9pm-6am	
Hours/time interval						Rarely/Never
Kitchen Oven						
Weekdays	6am-9am	9am-1pm	12-5pm	5pm-9pm	9pm-6am	
Hours/time interval						Rarely/Never
Weekend	6am-9am	9am-1pm	12-5pm	5pm-9pm	9pm-6am	
Hours/time interval						Rarely/Never
Dishwasher						
Weekdays	6am-9am	9am-1pm	12-5pm	5pm-9pm	9pm-6am	
Hours/time interval						Rarely/Never
Weekend	6am-9am	9am-1pm	12-5pm	5pm-9pm	9pm-6am	
Hours/time interval						Rarely/Never
Clothes Washer						
Weekdays	6am-9am	9am-1pm	12-5pm	5pm-9pm	9pm-6am	
Hours/time interval						Rarely/Never
Weekend	6am-9am	9am-1pm	12-5pm	5pm-9pm	9pm-6am	
Hours/time interval						Rarely/Never
Bathroom Fan						
Weekdays	6am-9am	9am-1pm	12-5pm	5pm-9pm	9pm-6am	
Hours/time interval						Rarely/Never
Weekend	6am-9am	9am-1pm	12-5pm	5pm-9pm	9pm-6am	
Hours/time interval						Rarely/Never
Clothes Dryer						
Weekdays	6am-9am	9am-1pm	12-5pm	5pm-9pm	9pm-6am	
Hours/time interval						Rarely/Never
Weekend	6am-9am	9am-1pm	12-5pm	5pm-9pm	9pm-6am	
Hours/time interval						Rarely/Never
Refrigerator						
Weekdays	Always			Never		
Weekends	Always			Never		
Other						
Weekdays	6am-9am	9am-1pm	12-5pm	5pm-9pm	9pm-6am	
Hours/time interval						Rarely/Never
Weekends	6am-9am	9am-1pm	12-5pm	5pm-9pm	9pm-6am	
Hours/time interval						Rarely/Never

7) What would you describe as the significant barriers that would prevent you from carrying out renovations to improve the energy efficiency of your home? Rank the following from 1 to 5 (5 being a major barrier and 1 being no issue)

Barrier	Ranking (1-5)
Cost	
Time Commitment	
Displacement from home	
Disruption (Noise, privacy, loss of functionality)	
Aesthetics (conservation of historic façade)	
Unclear Energy reduction results	

9. What energy savings goal would you hope to achieve through renovations?

_____
_____

10. Considering this savings goal what budget range would you consider reasonable (circle one)?
    a) $0-10,000   b) $10-20,000   c) $30-50,000   d) $50-75,000   e) $75-100,000

11. Are there any architectural features of your home that you would not consider changing through energy saving renovations? i.e. Windows, doors, brick façade.

Additionally are there any exterior facades that you would not consider over cladding to improve energy performance, if it meant minimizing disruption to interior spaces.

_____
_____
_____
_____

References:

[1] NRCAN, "Energy Efficiency Trends in Canada, 1990-2013: Residential sector," Natural Resources Canada, 2014.

[2] H. Karunathilake, K. Hewage, and R. Sadiq, "Opportunities and challenges in energy demand reduction for Canadian residential sector: A review," 2017.

[3] D. Jermyn and R. Richman, "A process for developing deep energy retrofit strategies for single-family housing typologies: Three Toronto case studies," *Energy Build.*, vol. 116, pp. 522–534, 2016.

[4] Natural Resources Canada, "Heating with Gas," 2012.

[5] "A Clearer View on Ontario's Emissions Electricity emissions factors and guidelines," 2019.

[6] B. Dong, C. Kennedy, and K. Pressnail, "Comparing life cycle implications of building retrofit and replacement options," 2005.

[7] K. M. Blaszak, R. Richman, and P. Eng, "Prioritizing Method for Retrofitting Toronto's Single-Family Housing Stock to Reduce Heating and Cooling Loads," 2013.

[8] "Neighbourhood Profiles – City of Toronto." [Online]. Available: https://www.toronto.ca/city-government/data-research-maps/neighbourhoods-communities/neighbourhood-profiles/. [Accessed: 29-Dec-2019].

[9] P. Westermann and Evins and R. Evins, "Surrogate modelling for sustainable building design-A review," in *Energy & Buildings*, 2019, vol. 198, pp. 170–186.

[10] D. Coakley, P. Raftery, and M. Keane, "A review of methods to match building energy simulation models to measured data," *Renew. Sustain. Energy Rev.*, vol. 37, pp. 123–141, 2014.

[11] A. I. J. Forrester, A. Sbester, and A. J. Keane, "Engineering Design via Surrogate Modelling: A Practical Guide - Intro," *Eng. Des. via Surrog. Model. A Pract. Guid.*, 2008.

[12] S. Seyedzadeh, I. Glesk, and M. Roper, "Machine learning for estimation of building energy consumption and performance: a review."

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

[14] M. Kavgic, A. Mavrogianni, D. Mumovic, A. Summerfield, Z. Stevanovic, and M. Djurovic-Petrovic, "A review of bottom-up building stock models for energy consumption in the residential sector," *Build. Environ.*, vol. 45, pp. 1683–1697.

[15] C. Cerezo, J. Sokol, S. AlKhaled, C. Reinhart, A. Al-Mumin, and A. Hajiah, "Comparison of

four building archetype characterization methods in urban building energy modeling (UBEM): A residential case study in Kuwait City," *Energy Build.*, vol. 154, pp. 321–334, 2017.

[16] L. G. Swan and V. I. Ugursal, "Modeling of end-use energy consumption in the residential sector: A review of modeling techniques," *Renew. Sustain. Energy Rev.*, vol. 13, no. 8, pp. 1819–1835, 2009.

[17] C. F. Reinhart and C. Cerezo Davila, "Urban building energy modeling - A review of a nascent field," *Building and Environment*, vol. 97. pp. 196–202, 2016.

[18] H. E. Zirnhelt and R. C. Richman, "The potential energy savings from residential passive solar design in Canada," *Energy Build.*, vol. 103, pp. 224–237, 2015.

[19] A. Mucciarone, "Towards a Proposed Framework for Analyzing Sustainable Renovation Building Envelope Assemblies," Ryerson University, 2011.

[20] S. Niger, "HIGH PERFORMANCE RETROFIT OPPORTUNITIES OF TORONTO ' S 1970S RESIDENTIAL DETACHED AND SEMI-DETACHED HOUSES By Sharmeen Niger Bachelor of Architecture , Ahsanullah University of Science & Technology , 2006 Master of Planning and Design , University of Melb," Ryerson University, 2016.

[21] L. G. Swan, V. I. Ugursal, and I. Beausoleil-Morrison, "Hybrid residential end-use energy and greenhouse gas emissions model-development and verification for Canada," 2011.

[22] L. G. Swan, V. I. Ugursal, and I. Beausoleil-Morrison, "A database of house descriptions representative of the Canadian housing stock for coupling to building energy performance simulation."

[23] S. Blais, M. Sc, A. Parekh, P. Eng, M. A. Sc, and L. Roux, "ENERGUIDE FOR HOUSES DATABASE-AN INNOVATIVE APPROACH TO TRACK RESIDENTIAL ENERGY EVALUATIONS AND MEASURE BENEFITS."

[24] "EnerGuide energy efficiency home evaluations | Natural Resources Canada." [Online]. Available: https://www.nrcan.gc.ca/energy-efficiency/energuide-canada/energuide-energy-efficiency-home-evaluations/20552. [Accessed: 26-Dec-2019].

[25] "ESP-r." [Online]. Available: http://www.esru.strath.ac.uk/Programs/ESP-r.htm. [Accessed: 01-Jan-2020].

[26] A. D. Wills, "On the modelling and analysis of converting existing Canadian residential communities to net-zero energy," 2018.

[27] "Energy Step Code 2017 Metrics Research - Full Report." [Online]. Available: https://www.bchousing.org/research-centre/library/residential-design-construction/energy-step-code-2017-full-report&sortType=sortByDate. [Accessed: 09-Dec-2019].

[28] D. Jermyn, "Deep energy retrofits: Toronto's urban single family housing stock. | Digital Repository," Ryerson University, 2014.

167

[29]   L. Swan, V. I. Ugursal, and I. Beasuoleil-Morrison, "IMPLEMENTATION OF A CANADIAN RESIDENTIAL ENERGY END-USE MODEL FOR ASSESSING NEW TECHNOLOGY IMPACTS."

[30]   E. Barnes, "BUILDING ENERGY SURROGATE MODELLING – A FEATURE SELECTION METHODOLOGY," Ryerson Univeristy, 2019.

[31]   D. Bzdok, N. Altman, and M. Krzywinski, "Points of Significance: Statistics versus machine learning," *Nature Methods*, vol. 15, no. 4. Nature Publishing Group, pp. 233–234, 03-Apr-2018.

[32]   J. S. Hygh, J. F. DeCarolis, D. B. Hill, and S. Ranji Ranjithan, "Multivariate regression as an energy assessment tool in early building design," *Build. Environ.*, vol. 57, pp. 165–175, 2012.

[33]   L. Magnier and F. Haghighat, "Multiobjective optimization of building design using TRNSYS simulations, genetic algorithm, and Artificial Neural Network," *Build. Environ.*, vol. 45, pp. 739–746.

[34]   P. G. Benardos and G. C. Vosniakos, "Optimizing feedforward artificial neural network architecture," *Eng. Appl. Artif. Intell.*, vol. 20, no. 3, pp. 365–382, Apr. 2007.

[35]   J. Moolayil, *Learn Keras for deep neural networks : a fast-track approach to modern deep learning with Python*. .

[36]   D. Barnes *et al.*, *Artificial Intelligence and Machine Learning Fundamentals*. 2018.

[37]   D. L. J. Alexander, A. Tropsha, and D. A. Winkler, "Beware of R2: Simple, Unambiguous Assessment of the Prediction Accuracy of QSAR and QSPR Models," *J. Chem. Inf. Model.*, vol. 55, no. 7, pp. 1316–1322, Jul. 2015.

[38]   T. Hastie, R. Tibshirani, G. James, and D. Witten, *An Introduction to Statistical Learning with Applications in R*, vol. 102. 2006.

[39]   J. F. Hair, W. C. Black, B. J. Babin, and R. E. Anderson, *Multivariate Data Analysis*. 2000.

[40]   D. Krstajic, L. J. Buturovic, D. E. Leahy, and S. Thomas, "Cross-validation pitfalls when selecting and assessing regression and classification models," 2014.

[41]   X. Chen, H. Yang, and K. Sun, "Developing a meta-model for sensitivity analyses and prediction of building performance for passively designed high-rise residential buildings," *Appl. Energy*, vol. 194, pp. 422–439, 2017.

[42]   S. A. R. Sangireddy, A. Bhatia, and V. Garg, "Development of a surrogate model by extracting top characteristic feature vectors for building energy prediction," *J. Build. Eng.*, vol. 23, pp. 38–52, 2019.

[43]   J. A. Wass, *Statistics in a Nutshell*, vol. 26, no. 1. O'Reilly Media, 2009.

[44]   C. F. Dormann *et al.*, "Collinearity: a review of methods to deal with it and a simulation study evaluating their performance," *Ecography (Cop.).*, vol. 36, no. 1, pp. 27–46, Jan. 2013.

[45] D. H. Vu, K. M. Muttaqi, and A. P. Agalgaonkar, "A variance inflation factor and backward elimination based robust regression model for forecasting monthly electricity demand using climatic variables," *Appl. Energy*, vol. 140, pp. 385–394, 2015.

[46] R. M. O'Brien, "A caution regarding rules of thumb for variance inflation factors," *Qual. Quant.*, vol. 41, no. 5, pp. 673–690, 2007.

[47] J. García, R. Salmerón, C. García, and M. del M. López Martín, "Standardization of Variables and Collinearity Diagnostic in Ridge Regression," *Int. Stat. Rev.*, vol. 84, no. 2, pp. 245–266, Aug. 2016.

[48] NCSS and LLC, "Ridge Regression."

[49] H.-X. Zhao, F. Magoulès, E. C. Paris, and G. Voie Des Vignes, "Feature Selection for Predicting Building Energy Consumption Based on Statistical Learning Method."

[50] H. Zou and T. Hastie, "Regularization and variable selection via the elastic net," *J. R. Stat. Soc. Ser. B (Statistical Methodol.*, vol. 67, no. 2, pp. 301–320, Apr. 2005.

[51] "About the Pocket - thepocket.ca." [Online]. Available: https://www.thepocket.ca/who-we-are/about-the-pocket/. [Accessed: 18-Dec-2019].

[52] "Google Maps." [Online]. Available: https://www.google.com/maps/@43.6647973,-79.346555,15z. [Accessed: 18-Dec-2019].

[53] "Weather Data | EnergyPlus." [Online]. Available: https://energyplus.net/weather. [Accessed: 20-Dec-2019].

[54] "EnergyPlus | EnergyPlus." [Online]. Available: https://energyplus.net/. [Accessed: 17-Dec-2019].

[55] "Welcome to Python.org." [Online]. Available: https://www.python.org/. [Accessed: 16-Dec-2019].

[56] "1. Supervised learning — scikit-learn 0.22 documentation." [Online]. Available: https://scikit-learn.org/stable/supervised_learning.html#supervised-learning. [Accessed: 16-Dec-2019].

[57] "tf.keras.wrappers.scikit_learn.KerasRegressor | TensorFlow Core r2.0." [Online]. Available: https://www.tensorflow.org/api_docs/python/tf/keras/wrappers/scikit_learn/KerasRegressor. [Accessed: 17-Dec-2019].

[58] "ARCHIVED - ecoENERGY Retrofit – Homes Program | Natural Resources Canada." [Online]. Available: https://www.nrcan.gc.ca/energy-efficiency/energy-efficiency-homes/what-energy-efficient-home/ecoenergy-retrofit-homes-program/5003. [Accessed: 26-Dec-2019].

[59] R. Qasass, M. Gorgolewski, and H. Ge, "Timber framing factors in Toronto residential house construction," *Archit. Sci. Rev.*, vol. 57, no. 3, pp. 159–168, 2014.

[60]   B.-L. Ahn, C.-Y. Jang, S.-B. Leigh, S. Yoo, and H. Jeong, "Effect of LED lighting on the cooling and heating loads in office buildings," *Appl. Energy*, vol. 113, pp. 1484–1489, 2014.

[61]   ASHRAE, *Standard 62.2-2019 -- Ventilation and Acceptable Indoor Air Quality in Residential Buildings*. 2019.

[62]   "Find & Compare Windows : Passive House Institute - United States." [Online]. Available: https://www.phius.org/phius-certification-for-buildings-products/phius-verified-window-performance-data-program/find-compare-windows. [Accessed: 17-Dec-2019].

[63]   "Office of Energy Efficiency | Natural Resources Canada." [Online]. Available: https://www.nrcan.gc.ca/energy/regulations-codes-standards/7049. [Accessed: 17-Dec-2019].

[64]   "MATLAB - MathWorks - MATLAB & Simulink." [Online]. Available: https://www.mathworks.com/products/matlab.html. [Accessed: 16-Dec-2019].

[65]   "OpenStudio | OpenStudio." [Online]. Available: https://www.openstudio.net/. [Accessed: 18-Dec-2019].

[66]   "3D Design Software | 3D Modeling on the Web | SketchUp." [Online]. Available: https://www.sketchup.com/. [Accessed: 18-Dec-2019].

[67]   "eppy.modeleditor — eppy 0.5.43 documentation." [Online]. Available: https://pythonhosted.org/eppy/_modules/eppy/modeleditor.html. [Accessed: 16-Dec-2019].

[68]   Q. Li and N. Lin, "The Bayesian Elastic Net," *Bayesian Anal.*, vol. 5, no. 1, pp. 151–170, 2010.

[69]   "TensorFlow." [Online]. Available: https://www.tensorflow.org/. [Accessed: 17-Dec-2019].

[70]   "Home - Keras Documentation." [Online]. Available: https://keras.io/. [Accessed: 16-Dec-2019].

[71]   D. Jermyn and R. Richman, "A process for developing deep energy retrofit strategies for single-family housing typologies: Three Toronto case studies," *Energy Build.*, vol. 116, pp. 522–534, 2016.

[72]   U.S. Department of Energy, "Residential Prototype Building Models," *U.S. Department of Energy*, 2017. [Online]. Available: https://www.energycodes.gov/development/residential/iecc_models. [Accessed: 03-Dec-2019].

[73]   N. Kruis, "Development and Application of a Numerical Framework for Improving Building Foundation Heat Transfer Calculations," 2015.

[74]   "CHREM/CSDDRD at master · ESP-rCommunity/CHREM · GitHub." [Online]. Available:

https://github.com/ESP-rCommunity/CHREM/tree/master/CSDDRD. [Accessed: 01-Jan-2020].