Title	Accurate identification of influential building parameters through an integration of global sensitivity and feature selection techniques
Authors(s)	Neale, John, Shamsi, Mohammad Haris, Mangina, Eleni, Finn, Donal, O'Donnell, James
Publication date	2022-06-01
Publication information	Neale, John, Mohammad Haris Shamsi, Eleni Mangina, Donal Finn, and James O'Donnell. "Accurate Identification of Influential Building Parameters through an Integration of Global Sensitivity and Feature Selection Techniques." Elsevier, June 1, 2022. https://doi.org/10.1016/j.apenergy.2022.118956.
Publisher	Elsevier
Item record/more information	http://hdl.handle.net/10197/26156
Publisher's version (DOI)	10.1016/j.apenergy.2022.118956

Accurate Identification of Influential Building Parameters through an integration of global sensitivity and feature selection techniques

John Neale[a,c], Mohammad Haris Shamsi[a,d], Eleni Mangina[b], Donal Finn[a], James O'Donnell[a,*]

[a]*School of Mechanical and Materials Engineering and UCD Energy Institute, University College Dublin, Belfield, Dublin 4, Ireland*
[b]*School of Computer Science and UCD Energy Institute, UCD, Dublin, Ireland*
[c]*Intel, Kildare, Ireland*
[d]*University of Victoria, BC, Canada*

Abstract

The development of building energy performance simulation models often requires significant time and effort to achieve an acceptable degree of prediction accuracy. As such, energy modelers introduce various simplifications and assumptions that require a high degree of modeling literacy to avoid any errors in energy predictions. Previous studies relate these simplifications to the identification of influential building parameters using engineering judgement techniques that are often subjective and differ based on experts' opinion. The proposed methodology accurately defines influential and non-influential building parameters to formulate a guideline minimum dataset in the context of residential building energy models. The methodology integrates two feature selection techniques (Bayesian Information Criteria and Least Absolute Shrinkage with Selection Operator) with parametric analysis to determine the set of influential parameters. The study uses Irish residential archetypes to compare and validate the subsets of influential parameters using sensitivity rankings and established validation metrics. The predicted annual energy use lies within 10% of measured data for both subsets of influential parameters. Thereby, energy modelers could significantly reduce the time and effort spent on model development while maintaining the desired accuracy. The formulated datasets represent only influential features and hence, could be used by urban planners and energy policymakers to estimate energy retrofit investment costs, emission reductions and energy savings.

Keywords: Energy modeling, Building Energy Performance Simulation, BEPS, Feature Selection, Sensitivity Analysis, Parametric Analysis

List of Abbreviations

ANOVA Analysis of Variance.

BEPS Building Energy Performance Simulation.

BIC Bayesian Information Criteria.

COP Coefficient of Performance.

CSO Central Statistics Office.

CV Cross Validation.

DSA Differential Sensitivity Analysis.

EPC Energy Performance Certificate.

EPSIH Energy Performance Survey of Irish Housing.

FS Feature Selection.

GSA Global Sensitivity Analysis.

INSHQ Irish National Survey of Housing Quality.

IWEC International Weather Energy Calculation.

LASSO Least Absolute Shrinkage and Selection Operator.

LSA Local Sensitivity Analysis.

MPE Mean Percentage Error.

PDF Probability Distribution Function.

RMSE Root Mean Square Error.

SA Sensitivity Analysis.

SEAI Sustainable Energy Authority of Ireland.

List of Symbols

μ_i^*	Absolute Mean of Parameter i
σ	Standard Deviation
a	Minimum Value
Aux	Auxiliary Energy Consumption (kWh)
b	Modal Value
c	Maximum Value
D	Variance
DHW	Domestic Hot Water Usage (m^3/s)
EE	Elementary Effects
$Equip$	Equipment Power Density (W/m^2)
$HSBT$	Heating Set-Back Temperature (°C)
$HSPT$	Heating Set-Point Temperature (°C)
IC	Influence Coefficient
IG	Information Gain
k	Parameter Identity
K_m	Conductivity value being modified (W/m^2K)
Ld	Normalized Lighting Density (lx)
N	Number of Samples
Occ	Occupant Density (people/m^2)
OP	Output Parameter
r	Number of trajectories
S_i	First Order Effect from Parameters i
S_{ij}	Second Order Effect from Parameters i and j
S_{Ti}	Total Order Effect from Parameters i
$SHGC$	Solar Heat Gain Coefficient

U_{S-S}	Surface-to-Surface U-Value (W/m^{2}K)
$UDoor$	Door U-Value (W/m^{2}K)
$UFloor$	Floor U-Value (W/m^{2}K)
$UFrame$	Frame U-Value (W/m^{2}K)
$UGlaze$	Glazed Portion U-Value (W/m^{2}K)
$UPart$	Internal Partition U-Value (W/m^{2}K)
$URoof$	Roof U-Value (W/m^{2}K)
$UWall$	External Wall U-Value (W/m^{2}K)
V_t	Light Transmittance Value (%)
WWR	Window-to-Wall Ratio (%)
X	N *k Matrix
$X^{1,2,..r}$	Trajectory point
X_i	Input Parameter i
Y	Model Output
Y_{Pred}	Vector of predicted values
Y_{Train}	Vector of observed values
Roof_abs	Roof Solar Absorption
Roof_E	Roof Emissivity
Wall_abs	Wall Solar Absorption
Wall_E	Wall Emissivity

1. Introduction

Despite recent technological advancements in energy technologies, building energy efficiency improvements have not resulted in any significant decline to the continually rising energy demand [1]. The final energy use of buildings grew from 118 EJ in 2010 to around 128 EJ in 2019. Direct and indirect emissions from electricity and commercial heat used in buildings rose to an all time high of 10 $GtCO_2$ in 2019. Approximately half of the building stock in the European Union is at least 50 years old with 75% of these buildings classified as ‘inefficient’ based on infrequent renovations [2]. According to a recent report by the European Commission, enhancing the existing energy efficiency through fabric enhancements and energy conversion system upgrades could reduce the current consumption trend by 5 to 6% with an associated 5% reduction in CO_2 emissions (2030 projections) [3].

The use of Building Energy Performance Simulation (BEPS) has become increasingly popular in the evaluation and assessment of building energy performance for new and existing buildings [4]. BEPS tools facilitate energy forecasting and energy performance analysis with various levels of detail [5]. Such tools implement a suite of detailed physical relations that employ numerous parameters, the details of which are often not available to modelers [6]. These parameters are set to default values based on previous literature, prior experience or industry standards that introduce a gap between simulated predicted and measured data [7]. Therefore, it is crucial to limit the number of parameters and associate a rank to each parameter in order to reduce the likelihood of incorrect interpretations. Using these limited parameters, modelers could have additional time to acquire accurate values of influential parameters [8]. Modelers could further quantify the uncertainty of the reduced set of parameters to build confidence in simulation results

Existing BEPS studies on building energy model formulation primarily focus on examining techniques which improve the post occupancy results of BEPS models [9]. While these techniques enhance the accuracy of post-occupancy BEPS models, the process provides little guidance as to how the pre-construction design model can be made more accurate [10]. A study by Shiel et al. formulated modeling guidelines to aid modelers in generating accurate models depending on the type of available data [11]. One of the significant modeling aspects identified in this study linked the overall accuracy of the model to modeler’s interpretation of the design performance of a building. Another study by Imam et al. investigated the role of modeling competency of the developer [12]. The authors identified the role of time limitations and the modeler’s ability to omit or include influential parameters with regard to the model output. Furthermore, assumptions and simplifications are unavoidable when developing Building Energy Performance Simulation (BEPS) models and the implication of these comes with experience [13].

Sensitivity Analysis (SA) is the main method for determining these sensitive parameters and has been widely investigated in numerous studies and across various sectors [14]. SA aids the identification of system boundaries in terms of influential parameters, prioritization of additional data collection, and verification and validation of a model [15]. When developing a BEPS model, it is important for the energy modeler to have a strong understanding of all parameters (influential and non-influential) within the model. Numerous studies implement SA to examine the characteristics of building thermal performance for various applications, such as building design and retrofit, energy model calibration, and climate change impact on buildings [16]. An identical procedure is followed when implementing SA for different applications; the fundamental difference in formulating SA lies in the definition of uncertainty or probability distribution of input factors [16].

Broadly categorized as local and global SA, the Local Sensitivity Analysis (LSA) operates on one parameter at a time while keeping other parameters constant at their reference value. LSA is particularly useful when formulating a BEPS model using limited resources due to the ability to outline the accurate set of influential parameters [17]. This technique considers an equal probability of occurrence for each parameter without considering the effect of range and shape of the Probability Distribution Function (PDF), thus, undermining any correlations between input parameters [18]. Differential Sensitivity Analysis (DSA), another variation of LSA, is commonly used in the BEPS domain to examine the relative influence of different input parameters [19]. Although DSA reduces the number of required simulations and hence the computation time, this method assumes the output varies linearly over the range of input values. Thus, DSA only imparts a partial view of the behavior of any BEPS model and most nonlinear models are a complex function of the multidimensional space of input factors [20].

LSA methods usually work as a preliminary elimination technique to reduce large sets of parameters to smaller, more manageable sets for Global Sensitivity Analysis (GSA). Heiselberg et al. used the one-parameter-at-a-time method and applied a "sensitivity index" to omit the less influential parameters prior to the application of GSA [21]. Methods that fall under this category work on a probabilistic framework whereby the effect and range of an input parameter are incorporated. Several methods exist under GSA including screening-based methods, regression methods and variance-based methods. The most popular GSA method in BEPS is the regression method due to the moderate computational cost and relative ease of implementation. Although SA methods identify the most influential parameters, many of these methods lack the means of separating influential parameters from the non-influential ones that would result in an accurately defined minimum data set. The majority of SA studies only provide an indication of the most influential parameters and fail to define a minimum dataset of influential parameters to produce accurate BEPS models. Furthermore, previous SA literature often defines the minimum dataset using the stepwise regression approach, which has been found to give unreliable results [22]. Although meta-models increase the reliability of regression approaches, their application is not set-up for defining a minimum dataset of parameters [23].

Feature Selection (FS) methods are a potential means of grouping only the most influential parameters. These methods identify a subset of input parameters that effectively describe the overall input data while diminishing the effects from noise or irrelevant parameters and still provide accurate prediction results within defined accuracy bounds [24]. FS methods comprise three categories, namely, filter, wrapper and embedded methods. Filter methods are used as a pre-processing technique to reduce the dimensionality experienced in wrapper methods, which employ a learning algorithm to score subsets of features according to their predictive value, thus evaluating the usefulness of a subset [25]. Embedded methods reduce the computational time of wrapper methods by including the FS method as part of the training process [26]. A limited number of studies exist in the literature that focus on techniques that identify the most important features required to enable simplified energy modeling of residential buildings. For instances, relevant features can be identified using energy simulation with statistical analysis [19]. Some of these variables include the building area, U-values, fuel type, dwelling type, and age band. Ali et al. implemented a hybrid FS process using a combination of engineering and data-driven approaches [27]. Using an available Energy Performance Certificate (EPC) database of approximately 1,000,000 residences, the authors identified a list of 16 influential parameters from an initial list of 203 parameters for the Irish residential building stock. Existing FS studies usually do not account for any interaction effects between the identified influential parameters.

This research aims to enhance the identification of influential parameters using an integration of sensitivity analysis and feature selection methods. The study does not aim to compare the relative accuracies of both techniques. The proposed methodology uses the global sensitivity analysis techniques to determine the sensitivity rankings of the input parameter data. Following this identification procedure, the methodology employs feature selection techniques on the influential parameter set to formulate a subset of influential parameters for the archetypes considered in this study.

The main objective of this study is to develop a systematic and integrated methodology that defines the minimum datasets of most influential parameters for BEPS models. The proposed methodology implements and compares the results of two GSA methods, namely, Morris and Sobol methods. These methods associate sensitivity rankings to each parameter to identify the list of influential parameters. The methodology further combines SA with FS techniques sequentially to validate the rankings associated with each parameter. This study compares two FS techniques, namely, Bayesian Information Criteria (BIC) and Least Absolute Shrinkage and Selection Operator (LASSO) regression. The set of most influential parameters identified using SA and FS form a guideline minimum set of accurately defined input data. This guideline dataset facilitates informed decisions regarding model assumptions and simplifications, which eventually increases the modeling accuracy.

The paper consists of the following sections: Section 2 describes the devised overarching methodology to identify minimum datasets of parameters influencing the building energy performance. Section 3 introduces a case study using the Irish residential archetypes to implement the devised methodology. Section 4 discusses the strengths, weaknesses and limitations of SA and FS methods. Section 5 describes the conclusion and future work.

2. Methodology

The increasing sophistication of BEPS has significantly extended the number of inputs required for a given model. Consideration of each parameter often requires appropriate modeling experience to handle the model assumptions. Furthermore, not all parameters in a BEPS model influence the outcome of building energy performance simulation analysis. This study proposes a generic and a sequential implementation of Sensitivity Analysis (SA) and Feature Selection (FS) techniques using Irish residential archetypes as the validation case study. These techniques have often been implemented individually and combined and a sequential formulation is still missing. The integration of these techniques not only enhances the identified parameter list qualitatively but also renders a generalized and structured list of parameters that could be used for different case studies. The originality of this research lies in the integration implementation of these techniques to identify a minimal dataset of influential parameters that provides sufficient model prediction accuracy for various applications. The generalized approach not only defines a guideline dataset for a given model but also provides a means of validating the set of identified influential parameters.

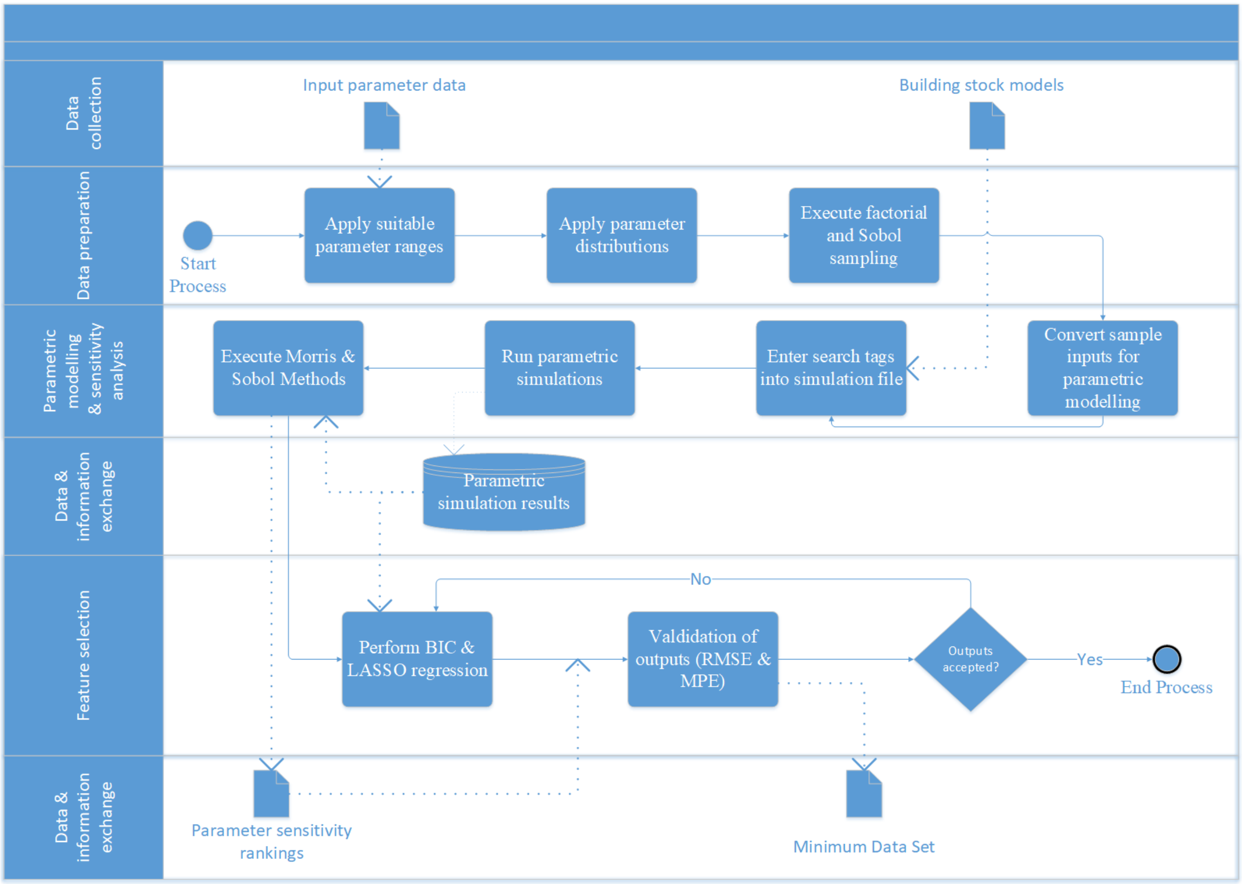

Figure 1: Process for identification of influential and non-influential parameters for a building.

The proposed methodology performs a SA on building parameters and implements FS algorithms to segregate the influential and non-influential parameters (Figure 1). The overall workflow comprises six processes, namely: 1) Data Collection, 2) Data Preparation, 3) Parametric Analysis, 4) Sensitivity Analysis, 5) Feature Selection and 6) Data Output and Exchange, to accurately identify influential and non-influential BEPS parameters.

2.1. Data Collection

Data collection is the initial process in the devised methodology and obtains BEPS models of the buildings under consideration. These BEPS models usually contain information to implement whole building energy analysis. Typical information in BEPS models include:

- Weather data representing the site;
- Building geometry, floor plan, construction materials, components and HVAC systems;
- Thermal zones inside the building;
- Occupancy patterns, lighting power loads, temperature setpoints and equipment operation by day, week and season and;

- Energy economics.

BEPS models are usually constructed ab initio after gathering the aforementioned data for design, renovation design or for use during building operation. Other BEPS application cases could include archetype modelling for a national building stock of any country (for instance, TABULA [28]).

2.2. Data Preparation

The data preparation process follows data collection and formulates the sample inputs for parametric analysis. This process consists of four sequential steps as follows:

1. Range Determination: Determine suitable ranges across which the parameters can vary. These ranges are extracted from the input parameter data;
2. Probability Distribution Identification: Assign probability distributions to describe the likelihood of the values within the selected range. This is an important step since the range and shape of the probability distribution attributed to each parameter can affect the overall SA results if non-linear interactions are present amongst the studied parameters;
3. Sample Generation: Generate samples from probability distributions to formulate the data for the implementation of SA and FS algorithms. This study demonstrates the application of Factorial and Sobol sampling techniques. The factorial sampling procedure converges faster when used in conjunction with the Morris method. Similarly, the Sobol sampling method converges on mean estimates when used when carrying out Sobol SA;
4. Input Conversion: Convert and arrange sample inputs for parametric analysis in line with the simulation environment (for instance, jEPlus [29]).

2.3. Parametric Analysis

Parametric analysis is executed prior to any sampling-based SA on a BEPS to generate samples of data for the input space of each parameter and the corresponding output. The degree of information given to the input parameters (described by the PDF) in a parametric analysis has a significant effect on the outcome of the SA [30, 31]. This process uses input samples generated in the data preparation process to execute parametric analysis using jEPlus [29].

jEPlus is an interface for EnergyPlus software to perform parametric simulations. The following three steps are crucial prior to the insertion of input samples in jEPlus.

1. Convert the simulation file into EnergyPlus IDF file format;
2. Enter Parameter Search Tags within these IDF files and;
3. Convert Input into Precursors for jEPlus/EnergyPlus. Certain parameters within the input sample matrix are converted into precursors, such as thermal conductivity, in order to be compatible with EnergyPlus.

After the execution of aforementioned steps, a parametric analysis is run in jEPlus using the converted input matrix to identify the output matrix.

2.4. Sensitivity Analysis

SA identifies the cluster of sensitive parameters and associates a sensitivity ranking to each parameter. This study implements two GSA techniques, namely Morris and Sobol methods.

The Morris SA method is a computationally inexpensive method widely used in the literature [32]. The Morris method is often combined with factorial sampling that generates parameter trajectories in order to efficiently cover the input space. One point within a trajectory represents one evaluation run of the model (i.e., one sample). The magnitude of variation in the model output, Y, due to variation in one parameter X is called elementary effect (EE) and is described as:

$$EE_i = \frac{Y(X + e_i \triangle_i) - Y(X)}{\triangle_i} \tag{1}$$

Where, X is a $N * k$ matrix of model inputs X with N samples and k parameters. e_i is a vector of zeros except for the ith components, which equals ±1 and represents the incremental change in parameters i.

The Sobol method falls under the group of Analysis of Variance (ANOVA) methods [33]. ANOVA SA is based on the functional decomposition scheme illustrated in Equation (2) that leads to the ANOVA decomposition of the total model variance [17]:

$$V(Y) = \sum_i V_i + \sum_i \sum_{j>i} V_{ij} + ... + V_{ij...k} \tag{2}$$

Where, $V_i = V(E(YX_i))$ is the influence each parameter has on the total variance of the output with the inner expectation operator denoting that the mean of Y is taken over all possible values of X_i, and $V_{ij} = V(E(YX_i, X_j)) - V_i - V_j$ is the interaction effect of parameters X_i and X_j. This process is continued for all higher order effects until $V_(ij...k)$. First and total order effects are normalized into indices between [0,1].

2.5. Feature Selection

Not all building parameters of a BEPS are equally likely to influence the building energy use. It is, therefore, crucial to identify these influential features that significantly influence building energy performance at individual building level. The FS process provides a potential means of grouping only the most influential parameters. This process removes irrelevant and redundant information and only selects the set of informative features that influence building energy performance. Amongst the different available FS methods, this study implements the embedded FS technique, Least Absolute Shrinkage and Selection Operator (LASSO), as this method combines the benefits of both filter and wrapper techniques (Figure 2). The computation time required for reclassifying different subsets in wrapper methods is reduced in embedded methods by incorporating the FS as part of the training process (Figure 2) [26].

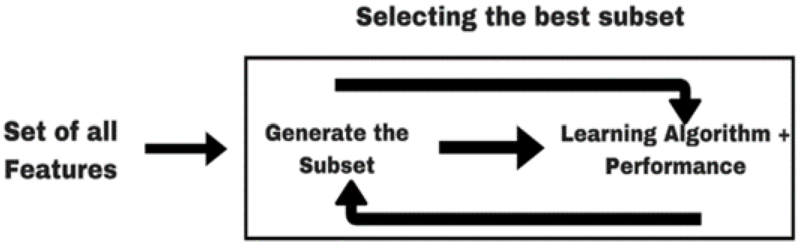

Figure 2: FS Process Chain Including Embedded Technique Illustrated in Component Block [34]

The LASSO technique was introduced by Tibshirani as a means of eliminating less informative variables in least squares multiple linear regression [35]. The Lagrangian form of the LASSO equation is defined by [36]:

$$\hat{\beta} = \min \sum_{i=1}^{N} \left(y_i - \beta_o - \sum_{j=1}^{p} x_{ij}\beta_j \right)^2 + \lambda \sum_{j=1}^{p} |\beta_j|_1 \tag{3}$$

Where, $\lambda \geq 0$ is a complexity parameter that controls the amount of shrinkage of the L_1-norm penalty. y_i is the response variable and x_{ij} is the covariate vector. This FS method is said to outperform other traditional filter and wrapper methods and has advantageous characteristics such as interpretability, generalization ability and numerical stability [37]. The important parameter in the LASSO equation is the regularization parameter λ, which determines the regression coefficients to be neglected, thus eliminating the respective parameters. Continuous shrinkage often improves the prediction accuracy due to the bias-variance trade-off. This study further integrates the Cross Validation (CV) approach to determine the value of λ. The CV approach takes the training data from the parametric analysis and splits it into k subsets and trains the regression model on $k - 1$ while holding out the last subset for validation.

The FS process further uses an exhaustive search algorithm capable of selecting an optimum subset model based on a certain information criteria. The main aim of this criteria is to determine a parsimonious model that gives a trade-off between bias (distance between the average estimate and truth) and variance (spread of the estimates around the truth). This study implements the Bayesian Information Criterion (BIC) to penalize model complexity and identify a model with lesser parameters than the original model (Equation 4).

$$BIC = K \cdot log(n) - 2\log(\varsigma) \tag{4}$$

Where, K is the number of estimable parameters (bias-correction term), $K \cdot \log(n)$ is penalty term and $\log(\varsigma)$ is the log-likelihood at its maximum point of the model estimated. It is worthwhile to mention that when analysing candidate models using BIC through scores and weights, a degree of ‘common sense’ is required to select the appropriate model. Questions concerning the strength of evidence for the models in the set are best addressed by doing an analysis of residuals, adjusted R^2, and other model diagnostics or descriptive statistics.

Finally, the FS results are evaluated using validation metrics such as Mean Percentage Error (MPE) and Root Mean Square Error (RMSE).

2.6. Data Output & Exchange

This is the last process of the proposed methodology and mostly organizes and classifies the obtained minimal output dataset to be used across different interoperable platforms.

3. Demonstration Cases

This study uses the Irish building archetypes that were originally developed by Neu et al. [38] and are based on a report by the Department of the Environment, Community and Local Government (DECLG) [39]. These are deemed to be representative of over 80% of the Irish building stock. The archetypes comprise of five geometric configurations; each with three different material sets (i.e., new insulated cavity wall, existing uninsulated cavity wall, and existing uninsulated hollow block wall), thus, resulting in fifteen BEPS models. The physical dimensions of the archetypes are consistent throughout each construction, therefore, these fifteen different models are reduced to five based on a similar research work by Famuyibo et al. [40]. The authors conducted a statistical analysis of two housing databases (Energy Performance Survey of Irish Housing (EPSIH) and the Irish National Survey of Housing Quality (INSHQ)), which led to the development of average, or "most probable" characteristics of different Irish dwelling archetypes. This justifies the use of five archetypes instead of fifteen. These five archetypes are used as the basis of this study and are as follows:

1. A two-storey detached dwelling (hereafter referred to as "detached");
2. A two-storey semi-detached dwelling (hereafter referred to as "semi-detached");
3. A single-storey detached dwelling (hereafter referred to as "bungalow");
4. A mid-floor apartment (hereafter referred to as "mid-flat") and;
5. A top-floor apartment (hereafter referred to as "top-flat").

Figure 3 offers a 3D representation of these archetypes. A more detailed description of these dwellings is included in Appendix A. Top-flat and mid-flat apartments are simulated by adjusting the boundary conditions of building surfaces. Table 1 enlists the 24 parameters included under the scope of this study. These parameters are extracted from the EPC datatset published by the Sustainable Energy Authority of Ireland (SEAI). The list of parameters includes building fabric parameters (U-values), building energy use, building equipment parameters (lighting densities), ventilation and occupancy.

3.1. Data Collection

The data collection process extracts BEPS data from various sources, namely, ASHRAE international weather database, the building census dataset and previous works published by [19, 38, 40]. The weather database comprises the International Weather Energy Calculation (IWEC) files for various locations in Ireland. These files are compatible with the

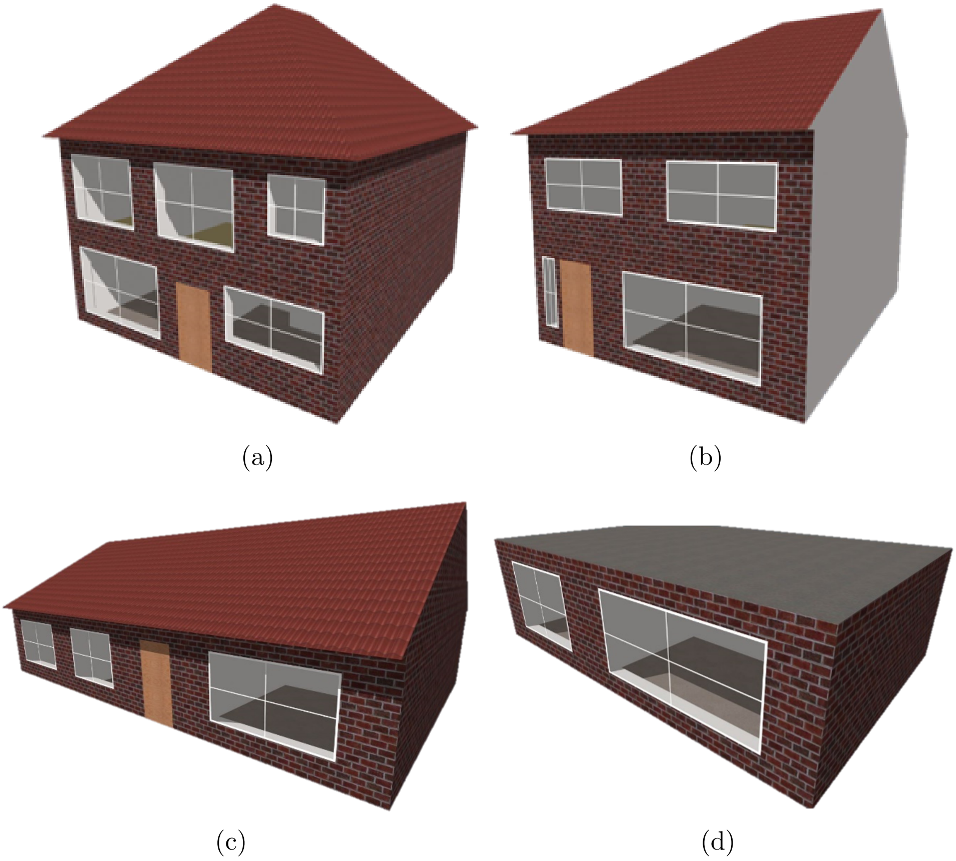
(a) (b) (c) (d)

Figure 3: 3D visualizations of model archetype dwellings illustrating a representative variation in archetype geometries: (a) Detached; (b) Semi-detached; (c) Bungalow; (d) Apartment.

EnergyPlus simulation engine. As a majority of the Irish residential dwellings are located in the Dublin county, this study simulates the archetypes using the EnergyPlus weather file of Dublin. The census dataset, published by the Central Statistics Office (CSO), provides the number of buildings in any area and other key information such as average person per household [41]. Previous studies provide the archetype BEPS models for simulation and these models contain information regarding building geometry, floor plan, construction materials, HVAC components, systems, thermal zones, occupancy schedules, lighting power loads, temperature setpoints and equipment operation [19, 38].

3.2. Data Preparation

The data preparation process extracts sample inputs for parametric analysis from the aforementioned data sources. The first step outlines the suitable range and modal values of building parameters across which the parameters can vary (see Table 1). The modal values represent the most common occurring parameter values in the Irish housing stock. It is worthwhile to mention that both the floor and roof U-values are not examined for the mid-floor apartment. This is due to the assumption the floor and ceiling are adjacent to

Table 1: Range & Modal Values of Building Parameters.

Parameter	Unit	Acronym	Min	Max	Bungalow	Detached	Semi-Detached	Top-Flat	Mid-Flat
Wall U-Value	W/m^2K	UWall	0.1	1.1	0.5	0.5	0.5	0.5	0.5
Roof U-Value	W/m^2K	URoof	0.1	0.9	0.33	0.33	0.33	0.33	0.33
Floor U-Value	W/m^2K	UFloor	0.1	1.1	0.58	0.58	0.5	N/A	N/A
Internal Partition U-Value	W/m^2K	UPart	1	3	2.075	2.075	2.075	2.075	2.075
Door U-Value	W/m^2K	UDoor	0.5	3	2.041	2.041	2.041	N/A	N/A
Frame U-Value	W/m^2K	UFrame	0.5	4.5	3.633	3.633	3.633	3.633	3.633
Glazed Portion U-Value	W/m^2K	UGlaze	0.6	4.6	3	3	3	3.25	3.25
Window-to- Wall Ratio	%	WWR	10	70	24.8	22.23	24.41	30.5	30.5
Solar Heat Gain Coefficient		SHGC	0.1	0.9	0.4	0.4	0.4	0.4	0.4
Light Transmittance Value		Vt	0.19	0.99	0.56	0.56	0.56	0.56	0.56
Domestic Hot Water Usage	L/m^2day	DHW	0.5	3.5	158.9	158.9	110.5	85.6	85.6
Auxiliary Energy Consumption	kWh/m^2yr	Aux	1	5	2.13	1.75	1.89	2.98	3.19
Heating System Seasonal COP		COP	0.5	2.5	0.8	0.8	0.8	0.8	0.8
Building Orientation	°	Orientation	0	359	N/A	N/A	N/A	N/A	N/A
Normalized Lighting Density	W/m^2-100lx	Ld	1	9	3.025	2.92	2.95	3.29	3.29
Equipment Power Density	W/m^2	Equip	1	21	1.56	1.47	1.61	1.9	1.9
Wall Solar Absorptivity		Wall_abs	0.15	0.95	0.7	0.7	0.7	0.7	0.7
Roof Solar Absorptivity		Roof_abs	0.15	0.95	0.7	0.7	0.7	0.85	N/A
Wall Emissivity		Wall_E	0.15	0.95	0.9	0.9	0.9	0.9	0.9
Roof Emissivity		Roof_E	0.15	0.95	0.9	0.9	0.9	0.9	N/A
Heating Set-Point Temperature	$°C$	HSPT	18	23	21	21	21	21	21
Heating Set-Back Temperature	$°C$	HSBT	10	14	12	12	12	12	12
Air Changes Per Hour	h^{-1}	ACR	0.5	1.5	0.74	0.87	0.94	0.87	0.87
Occupancy	Person(s)	Occ	0	7	2.7	2.7	2.7	2.7	2.7

the apartments above and below with similar internal temperatures. Therefore, these are assumed to be adiabatic surfaces. The same assumption is made for the top-floor apartment except the roof U-value is considered. Finally, the external door for both the mid and top-floor apartment is assumed to open on to a corridor of similar internal temperature, and therefore represents adiabatic surface.

The second step is to assign probability distributions to each parameter, which requires an extensive size of the parameter input data. The probability distributions are assigned and generated using the SimLab v2.2 SA software [42]. The EPC database published by the SEAI offers a large database of several building energy performance parameters, however, this database only covers four of the twenty-four parameters being studied in this work. Due to this lack of readily available parameter information, this study employs a triangular distribution to describe the parameter input space. This approach is deemed suitable given that a triangular distribution offers an appropriate representation to describe the probability of a parameter value when there is insufficient data available. It is worthwhile to mention that this study assigns a uniform distribution to the building orientation parameter. while building designers look to optimize solar gains by placing the majority of glazed windows to the south, the orientation of the building is typically dependant on the direction of the adjacent roadway. Previous studies often do not define the average building orientation and given the vast variations in roadway directions, the parameter value is deemed equally probable.

The third step in the data preparation process involves input sampling using factorial and Sobol sampling methods. The factorial sampling method is essential to execute Morris SA and involves a number of individually randomized one-factor-at-a-time samples, where all parameters are varied in such a way that covers the entire input space, thus, giving global characteristics to the Morris method. The BEPS model is represented by a function y(x) where y is the output variable of interest (scalar) and X is a vector of real input variables with k-coordinates, each input variable being defined within the range of a continuous interval. Factorial sampling creates several trajectory points of design parameters until all the parameters are characterized by two different values, therefore, creating a set of (k + 1) independent parameter vectors which closes the trajectory. This procedure is repeated r times creating a set of r(k+1) trajectories [21]. Another factor analyzed is the number of levels (i.e. p-value). The p-value is strictly linked to the choice of r and should attain a high value for a high value of r value is chosen for better exploration of the input space [43].

Table 2: Number of model runs to generate Sobol samples.

Building Archetype	No. of Parameters	No. of Model Runs
Bungalow	24	102,400
Detached	24	102,400
Semi-Detached	24	102,400
Top-Floor Apartment	22	94,208
Mid-Floor Apartment	19	81,920

The third step further implements the Sobol sampling technique to implement the Sobol SA. Based on Sobol sequences, this technique has high convergence capabilities when performing SA. Sobol sequences are basically samples of multiple parameters uniformly distributed over a multi-dimensional parameter space. An important consideration when using this sampling method is to ensure that there are a sufficient number of model evaluations (i.e. sample size) to ensure convergence of the sensitivity indices [44]. The computational cost associated with the Sobol method is N(2k+2) and the maximum allowable samples in SimLab is 2,048. Table 2 indicates the number of model runs required for all five building archetypes. The total computational cost required to generate these samples is approximately 300 hours (Intel core i7, 2.80 GHz, 16 GB RAM).

Prior to the export of generated samples from SimLab to jEPlus, the fourth step modifies certain parameters to ensure their compatibility with EnergyPlus. EnergyPlus simulation engine does not offer an input space to enter the U-values for certain materials except for glazed window U-values. This step formulates precursors for the following materials in order to obtain the required change in U-value. The U-value of window frames also require modifications to ensure compatibility with EnergyPlus. To alter the WWR, this study varies the bottom two vertices of each window while holding the the top two vertices constant.

1. External wall U-value (Precursor: Thermal Conductivity);
2. Roof U-value (Precursor: Thermal Conductivity)
3. Floor U-value (Precursor: Thermal Conductivity)
4. Internal Partition U-Value (Precursor: R-Value)
5. External Door U-Value (Precursor: R-Value)

Table 3: Occupant Density Ranges & Modes for jEPlus (people/m^2).

Archetype	Unit	Min	Mode	Max
Bungalow	people/m^2	0	0.163	0.421
Detached	people/m^2	0	0.173	0.447
Semi-Detached	people/m^2	0	0.141	0.367
Top-Flat	people/m^2	0	0.206	0.535
Mid-Flat	people/m^2	0	0.206	0.535

Table 4: Conversion of Normalized Lighting Density to Lighting Density for the Bungalow.

		Normalized Lighting Density (W/m^2 – 100 lux)			Lighting Density (W/m^2)		
Zone	Required Lux	Lower	Modal	Upper	Lower	Modal	Upper
Kitchen	300	1	3.025	9	3	9.08	27
Corridor	100	1	3.025	9	1	3.03	9
Dining	100	1	3.025	9	1	3.03	9
Living Room	150	1	3.025	9	1.5	4.54	13.5
Bathroom	150	1	3.025	9	1.5	4.54	13.5
Bed Room 1	100	1	3.025	9	1	3.03	9
Bed Room 2	100	1	3.025	9	1	3.03	9
Bed Room 3	150	1	3.025	9	1.5	4.54	13.5
		Average (Exc. Kitchen)			1.214	3.67	10.93

The occupant density is represented as people/m^2 on a per zone basis in order to input the number of people into EnergyPlus and DesignBuilder (Table 3). To determine the range and mode for each archetype, this study uses the zone with the largest floor area to ensure that the maximum number of people does not exceed 7 (i.e. the upper limit). The Domestic Hot Water (DHW) parameter is transformed using Equation 5.

$$DHW(m^3/s) = \frac{DHW(l/m^2 - day) \times 0.001 \times A}{Hours/day \times 3600} \tag{5}$$

Where, A is the area of the zone. Hours/day is the number of hours based on the occupant schedule in that zone. For simplicity, DHW only serves the kitchen zone in each archetype. Hours/Day is set to 3.8 for all archetypes since the kitchen schedule is consistent across all models.

The input normalized lighting density is converted to lighting density using normalized lighting density based on the lighting density (W/m^2) and the required lux (Table 4). The average lighting density represents the density value of each zone except the kitchen. Including the lighting density of kitchen would strongly increase the average value.

3.3. Parametric Analysis

The parametric analysis process simulates each input parameter at r_i points within the identified parameter range to formulate how the output Y varies as a function of each input parameter X_i. This provides a sufficient resolution of how annual building energy consumption varies as a function of each input parameter examined. This study uses jEPlus to execute parametric simulations for each of the five Irish residential archetypes. These simulations provide the parametric data to implement SA using the Morris and Sobol methods.

3.4. Sensitivity Analysis

SA identifies the cluster of influential parameters. This study implements and compares the results of two GSA techniques to allow for a better validation of FS results. The first GSA technique, Morris method, is a computationally inexpensive method and is widely used in the literature. The Morris method is combined with factorial sampling in order to efficiently cover the input space. The second GSA technique, Sobol method, employs ANOVA analysis and uses Sobol sampling to generate the input space.

3.4.1. Morris Sensitivity Rankings

The Morris method uses the trajectories of design parameters generated by factorial sampling. One point within a trajectory represents one evaluation run of the model (i.e. one sample). For this study, 200-250 samples are produced depending on the building archetype. The sample size is based on the chosen values for r (set of trajectories) and p. A value of r = 10 facilitates the evaluation of 10 EE per parameter. The statistical measures used in SimLab for evaluating EEs include absolute mean μ^* and standard deviation σ.

Table 5: Morris results' classification scheme suggested by [14].

Impact on Output	Characterization Range
Non-linear and/or non-monotonic	$k > \sigma/\mu^* = 1$
Almost monotonic	$\sigma/\mu^* = 1\ k > \sigma/\mu^* = 0.5$
Monotonic	$\sigma/\mu^* = 0.5 > k > \sigma/\mu^* = 0.1$
Linear	$\sigma/\mu^* = 0.1 > k$

For each SA result hereafter, each point in the plot represents the average absolute mean values μ^* (x-axis) of the 10 trajectories and the standard deviations σ (y-axis) of the EE for each parameter. Also, the EE maintains the same physical unit as the model output, which is annual energy consumption (kWh/year) and annual heating load (kWh/year). Therefore, the points in each plot represents the magnitude of variation in annual energy consumption and heating demand when modifying one input parameter at a time. Additionally, to offer further insight in to the behaviour of input parameters, this study uses a classification scheme whereby the ratio of σ/μ^* indicates the impact of input parameters (k) in terms of non-linearity and non-monotonicity (Table 5). Building heating load signifies the heating demand (space heat energy required to maintain a set temperature) of the building and building electrical load signifies the electricity demand (electrical energy consumed by building equipment) of the building.

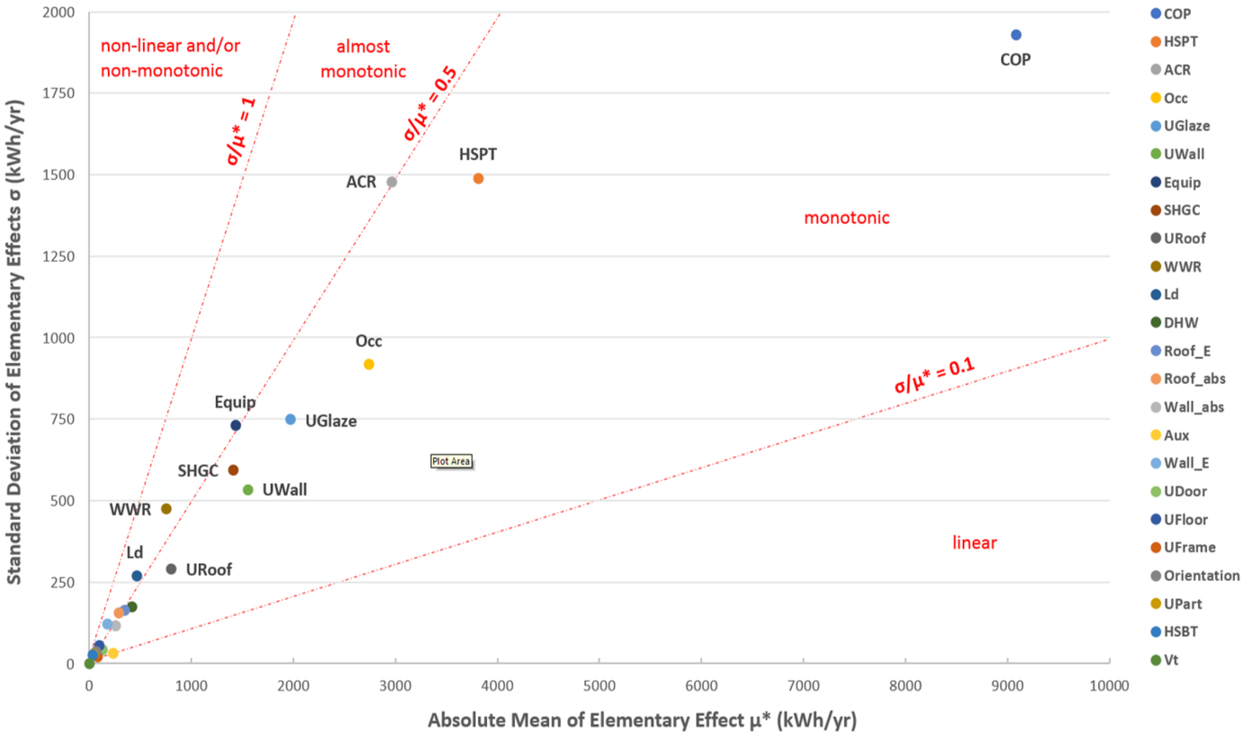

Figure 4: Morris Methods Results for Bungalow Energy Consumption.

Morris SA Results for the Bungalow Archetype: When considering the building energy consumption for the Bungalow archetype, the heating system seasonal Coefficient of Performance (COP) is the most influential parameter as the influence of COP is considered following the impact of remaining parameters (Figure 4). Also, since COP ranges from 0.8 (for boilers) to 2.5 (for heat pumps), a substantial effect on the output variance would have been expected. A majority of the input parameters have a monotonic or near monotonic relationship with the model output where only the auxiliary energy parameter exhibits a linear relationship.

A notable observation is that there does not appear to be any interaction effects between parameters (i.e. no parameter $> \sigma/\mu^* = 1$). This is somewhat surprising since a relationship would have been expected between parameters such as WWR and orientation or SHGC. Also, WWR is not entirely monotonically related to energy consumption which perhaps indicates that the interaction effect is not significant enough to alter this relationship to a non-linear and/or non-monotonic correlation. It is worthwhile to note that occupant density (Occ) lies in the top 10 ranks of the influential parameters which coincides with the findings of other SA studies, thereby, establishing the influence of occupancy level on building energy consumption [9].

The visible transmittance (V_t) value of windows has zero impact on energy consumption (Figure 4). This can be attributed to the fact that the lighting templates used in the DesignBuilder models are operated on an on/off schedule and are not governed by outdoor luminance levels. Although this is not ideal, the V_t values of windows should have an effect on lighting energy requirements. This will require further investigation and development of a method to test the effect of window V_t values on zones in which lighting levels are not controlled by the level of daylight.

Table 6: Bungalow Morris Rankings for Energy Consumption, Heating & Electrical Load.

Bungalow Morris Rankings			
Rank	Energy	Heating Load	Electrical Load
1	COP	HSPT	Equip
2	HSPT	Occ	Ld
3	ACR	ACR	Aux
4	Occ	UGlaze	-
5	UGlaze	Equip	-
6	UWall	SHGC	-
7	Equip	UWall	-
8	SHGC	URoof	-
9	URoof	WWR	-
10	WWR	Ld	-
11	Ld	DHW	-
12	DHW	Roof_abs	-
13	Roof_E	Roof_E	-
14	Roof_abs	Wall_abs	-
15	Wall_abs	Wall_E	-
16	Aux	UDoor	-
17	Wall_E	UFloor	-
18	UDoor	UFrame	-
19	UFloor	Orientation	-
20	UFrame	UPart	-
21	Orientation	HSBT	-
22	UPart	Vt	-
23	HSBT	-	-
24	Vt	-	-

For the Bungalow heating load parameters, the HSPT is the most significant parameter (Figure 5). Although there was a change in the ranking of ACR and occupant density, the relative difference in magnitude of these parameters is found to be insignificant similar to the SA results for energy consumption (Figure 4) The Morris method is known be biased by the occurrence (or absence) of outliers giving some variations in results with each run [17]. Therefore, the change in ranking could be due to slight differences in values selected during the development of trajectories in SimLab (Table 6).

The SA further considers the entire input parameter space to determine the electrical load rankings, however, as expected, only equipment density, lighting density and auxiliary energy exhibit any influence on the model outcome. The relative magnitudes between each parameter is also significant with equipment density ($\mu^* = 3230$, $\sigma = 345$) being approximately 3 times that of lighting density ($\mu^* = 1120$, $\sigma = 148$) and 14 times of auxiliary energy ($\mu^* = 230$, $\sigma = 33$). While auxiliary energy does not appear to be highly sensitive to the electrical load and energy consumption, this parameter is known to be responsible for significant proportions of CO_2 emissions associated with HVAC systems [45]. The Morris SA results for the remaining archetypes are included in Appendix B.

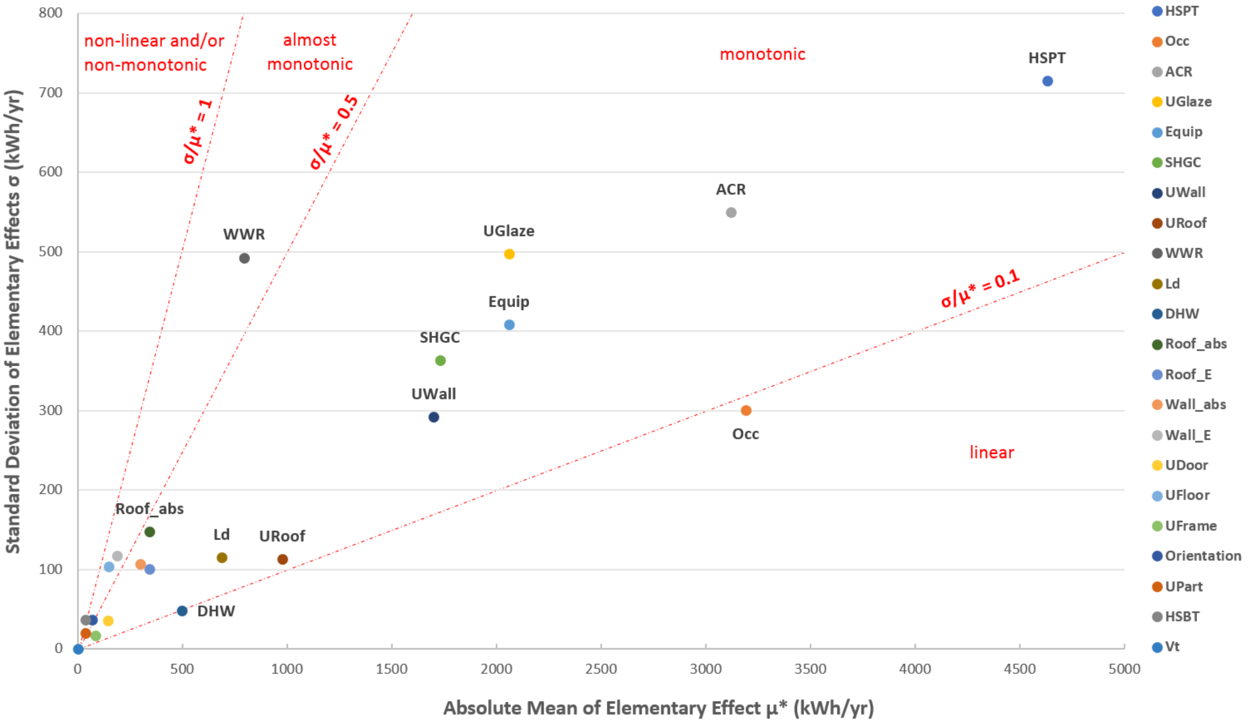

Figure 5: Morris Methods Results for Bungalow Heating Load.

3.4.2. Sobol Method

The Sobol method uses ANOVA decomposition to identify the influence and interaction of each parameter on the total variance of the output. The interaction defines the effect of different parameters in varying degrees of order. Given the significant computational expense to determine the second and higher order effects, this study only focuses on the first order effects that provide sufficient information on input parameters. Also, as this SA measure evaluates the parameter rankings to validate the FS results, the additional information on higher order effects is somewhat unnecessary. The Sobol sensitivity indices are calculated using the SA function in SimLab. The convergence of indices requires a large number of samples, which is addressed by selecting the largest available samples allowed in SimLab. However, for parameters with small indices, there is potentially some inherent inaccuracy. All first order indices are omitted for electrical load since there are no observed interaction effects (i.e. no change between first and total order indices).

Table 7: Bungalow Sobol Sensitivity Indices (%)

Energy Consumption (%)			Heating Load (%)			Electrical Load (%)	
1st Order Effect	Total Order Effect	Rank	1st Order Effect	Total Order Effect	Rank	Total Order Effect	Rank
65.3227	69.2399	COP	42.2027	42.8718	HSPT	90.31	Equip
13.6982	15.4919	HSPT	17.4590	17.6815	ACR	9.63	Ld
5.7758	6.7240	ACR	9.9956	10.1570	Equip	0.36	Aux
2.5071	2.7941	Occ	8.0326	8.0715	Occ	0	COP
2.3736	2.7841	UGlaze	6.7506	7.6831	UGlaze	0	HSPT
1.9613	2.2125	Equip	5.2616	5.5399	UWall	0	ACR
1.8517	2.0746	UWall	4.1263	4.4864	SHGC	0	Occ
1.3195	1.7018	SHGC	0.4147	1.6313	WWR	0	UGlaze
0.2096	0.7158	WWR	1.4684	1.5124	URoof	0	UWall
0.3990	0.4105	URoof	0.8939	0.9180	Ld	0	SHGC
0.2090	0.2752	Ld	0.4962	0.5062	DHW	0	WWR
0.1695	0.1748	DHW	0.2145	0.2907	Wall_abs	0	URoof
0.1086	0.1412	Wall_abs	0.1889	0.2195	Roof_abs	0	DHW
0.0842	0.0883	Roof_E	0.1594	0.1632	Roof_E	0	Wall_abs
0.0602	0.0770	Roof_abs	0.1300	0.1423	Wall_E	0	Roof_E
0.0363	0.0668	Wall_E	0.0585	0.0617	UDoor	0	Roof_abs
0.0449	0.0449	Aux	0.0254	0.0268	UFloor	0	Wall_E
0.0159	0.0184	UDoor	0.0068	0.0219	HSBT	0	UDoor
0.0123	0.0179	UFloor	0.0042	0.0170	Orientation	0	UFloor
0.0068	0.0070	UFrame	0.0151	0.0158	UFrame	0	UFrame
0.0017	0.0038	Orientation	0.0038	0.0063	UPart	0	Orientation
0.0015	0.0036	UPart	0	0	Vt	0	UPart
0.0006	0.0006	HSBT	0	0	COP	0	HSBT
0	0	Vt	0	0	Aux	0	Vt

Sobol Results for the Bungalow Archetype: The heating system seasonal COP exhibits a substantial influence on the output variance of the bungalow energy consumption (Table 7). The interaction effect of COP with other input parameters accounts for approximately 4% of the output variance in energy consumption. This interaction effect is somewhat expected given that the effect of other input parameters is taken into account prior to the calculation of the COP. Therefore, any parameter that increases the heating load will exemplify the effect of the heating system COP on the energy consumption. All the parameters possess some interaction effect. However, the overall effect of certain parameters is highly dependent on their interaction effect with other parameters. For instance, the interaction with other parameters significantly influences the WWR. The higher order effects of WWR only account for 0.51% and 1.22% of the output variance for the buildings energy consumption and heating load. While these higher order effects are somewhat insignificant, these account for more than three times of the overall effect of WWR on energy consumption and heating load outcome.

The Sobol SA results for the remaining archetypes are discussed in Appendix C.

3.4.3. *Comparison between Morris & Sobol Rankings*

Table 8: Comparison Between Energy Consumption Rankings from Morris & Sobol Methods (Green text Indicates Direct Match)

Bungalow		Detached		Semi-Detached		Top-Floor Apartment		Mid-Floor Apartment	
Morris	Sobol	Morris	Sobol	Morris	Sobol	Morris	Sobol	Morris	Sobol
COP	COP	COP	COP	COP	COP	COP	COP	COP	COP
HSPT	HSPT	HSPT	HSPT	HSPT	HSPT	ACR	ACR	ACR	ACR
ACR	ACR	Occ	Occ	Occ	ACR	Occ	HSPT	Occ	Occ
Occ	Occ	UWall	ACR	ACR	Occ	HSPT	Occ	HSPT	HSPT
UGlaze	UGlaze	ACR	UWall	UGlaze	UGlaze	URoof	URoof	UGlaze	Equip
UWall	Equip	Equip	Equip	Equip	Equip	Orientation	Equip	Equip	UGlaze
Equip	UWall	UGlaze	UGlaze	UWall	SHGC	Equip	UGlaze	UWall	SHGC
SHGC	SHGC	SHGC	SHGC	SHGC	UWall	UGlaze	Orientation	SHGC	Orientation
URoof	WWR	Ld	WWR	Ld	WWR	SHGC	SHGC	Orientation	UWall
WWR	URoof	URoof	URoof	WWR	Ld	DHW	DHW	DHW	DHW
Ld	Ld	WWR	Ld	URoof	URoof	UWall	UWall	WWR	WWR
DHW	DHW	DHW	Wall_abs	DHW	Wall_abs	WWR	Ld	Ld	Ld
Roof_E	Wall_abs	Wall_abs	DHW	Orientation	Roof_abs	Ld	WWR	Aux	Aux
Roof_abs	Roof_E	Aux	Wall_E	Wall_abs	DHW	Roof_abs	Roof_E	Wall_abs	Wall_abs
Wall_abs	Roof_abs	Roof_abs	Aux	Aux	Orientation	Roof_E	Roof_abs	Wall_E	Wall_E
Aux	Wall_E	Roof_E	Roof_abs	Roof_abs	Aux	Aux	Aux	UFrame	HSBT
Wall_E	Aux	Wall_E	UFrame	Roof_E	Roof_E	Wall_abs	Wall_abs	HSBT	UFrame
UDoor	UDoor	UDoor	Roof_E	Wall_E	UFloor	Wall_E	Wall_E	UPart	UPart
UFloor	UFloor	UFloor	UFloor	UFloor	Wall_E	UFrame	HSBT	Vt	Vt
UFrame	UFrame	UFrame	UDoor	UDoor	UFrame	HSBT	UFrame	-	-
Orientation	Orientation	Orientation	Orientation	HSBT	UDoor	UPart	UPart	-	-
UPart	UPart	UPart	HSBT	UFrame	HSBT	Vt	Vt	-	-
HSBT	HSBT	HSBT	UPart	UPart	UPart	-	-	-	-
Vt	Vt	Vt	Vt	Vt	Vt	-	-	-	-

The COP of the heating system is the most influential parameter as identified using both techniques. The results from two GSA methods in this study closely match one another as shown by the green text (Table 8). The top 10 parameters match exactly for each of the five archetypes. There are a few discrepancies between certain parameters, however, the differences are marginal with a strong majority of parameters differing only by ±1 in ranking and at most by ±3. The reason for these inconsistencies could be due to the Morris method's susceptibility to outliers, thus, introducing some variations in results with each run [17]. Another reason could be the lack of convergence on the Sobol indices, however, this tends to be problematic only for parameters exhibiting an insignificant amount of variance. Apart from the aforementioned issues and any inherent errors associated with the software, the assumptions considered in this might result in observed difference in rankings.

3.5. *Feature Selection*

The FS process identifies the influential and non-influential parameters influencing BEPS. This study implements two FS techniques as a measure of validating SA results. The first technique implements an exhaustive search model selection algorithm using BIC that determines the effectiveness of the model selection criteria. The second technique involves regularized logistic regression models known as LASSO regression. Regularization is frequently used in the field of machine learning to limit model complexity and avoid overfitting.

3.5.1. Exhaustive Search Model Selection via BIC

The BIC is used in conjunction with an exhaustive search algorithm to assess all possible model subsets. BIC is useful for generating smaller parameter subsets as too many parameters in a model may lead to overfitting. While the Bayesian criteria is used to evaluate the model subsets, an exhaustive search algorithm is used to generate the various parameter combinations to develop these model subsets. This study uses the R statistical software package MuMIn developed by Barton to formulate all possible model subsets and subsequently, rank each subset using BIC [46]. This model selection method uses the data generated by the Sobol sampling technique, which offers an extensive description of the parameters' input space. Prior to the application of BIC, the dataset is split into train and test data. While there is no defined correct split, a commonly used rule-of-thumb in the field of machine learning is to apply a 50/50 split (i.e. 50% of the data for training, 50% for testing). The results outline the top models selected from the exhaustive search algorithm using BIC and include various metrics for determining the optimum model subset. All the parameters are listed according to their Sobol rankings.

BIC Results for the Bungalow Archetype: When considering the energy consumption, the BIC scores indicate that Model 1 is the most accurate subset of the original set of parameters (Table 9). The weight metric also heavily supports Model 1 compared to Models 2 and 3. Using the BIC differences (△BIC) guide from Burnham and Anderson [47], the empirical support for Models 2 and 3 is considerably less than Model 1. Model 3 further penalizes the complexity and results in 16 parameters rather than 17. Interestingly, the top 3 BIC heating load models exclude the building orientation parameter. However, the BIC energy consumption model 3 only excludes this parameter. This potentially suggests that Model 3 is a better fit than Model 1 when considering the BIC results for energy consumption. Also, Model 3 conforms better with the Sobol rankings. Considering this degree of uncertainty, Model 3 is chosen as the final BIC model for this archetype.

The metrics for evaluating the BIC heating load results are not as definitive as the energy consumption results (Table 9). While the BIC score is in favour of Model A, the weight and △BIC do not significantly favour Model A over Model B. However, when observing the parameters within Model B, the analysis includes the internal partition U-value in the list of influential parameters. Based on engineering judgement, the internal partition U-value usually has negligible effects on energy consumption and heating demand. This assumption is confirmed in both SA tests carried out in this study with UPart being ranked as either the second or third least influential parameter in each archetype. As model A involves fewer parameters and excludes UPart from the list, this model is chosen as the final heating load BIC model for this archetype.

3.5.2. LASSO Method

The LASSO method identifies influential parameters by minimizing the residual sum of squares subjected to the condition that sum of the absolute value of the coefficients stays less than a constant. This study uses R statistical software to apply the LASSO regression and CV to each building archetype. The LASSO regression uses the glmnet package developed by Friedman et al. [48]. The CV approach uses the cv.glmnet function from the same R

Table 9: Bungalow Archetype BIC Results for Energy Consumption and Heating Demand.

Energy Consumption BIC Results				Heating Load BIC Results			
Parameter	Model 1	Model 2	Model 3	Parameter	Model A	Model B	Model C
COP	✓	✓	✓	HSPT	✓	✓	✓
HSPT	✓	✓	✓	ACR	✓	✓	✓
ACR	✓	✓	✓	Equip	✓	✓	✓
Occ	✓	✓	✓	Occ	✓	✓	✓
UGlaze	✓	✓	✓	UGlaze	✓	✓	✓
Equip	✓	✓	✓	UWall	✓	✓	✓
UWall	✓	✓	✓	SHGC	✓	✓	✓
SHGC	✓	✓	✓	WWR	✓	✓	✓
WWR	✓	✓	✓	URoof	✓	✓	✓
URoof	✓	✓	✓	Ld	✓	✓	✓
Ld	✓	✓	✓	DHW	✓	✓	✓
DHW	✓	✓	✓	Wall_abs	✓	✓	✓
Wall_abs	✓	✓	✓	Roof_abs	✓	✓	✓
Roof_E	✗	✗	✗	Roof_E	✓	✓	✓
Roof_abs	✓	✓	✓	Wall_E	✓	✓	✓
Wall_E	✓	✓	✓	UDoor	✓	✓	✓
Aux	✓	✓	✓	UFloor	✗	✗	✗
UDoor	✗	✗	✗	HSBT	✗	✗	✗
UFloor	✗	✓	✗	Orientation	✗	✗	✗
UFrame	✗	✗	✗	UFrame	✗	✗	✗
Orientation	✓	✓	✗	UPart	✗	✓	✗
UPart	✗	✗	✗	Vt	✗	✗	✓
HSBT	✗	✗	✗				
Vt	✗	✗	✗				
BIC Score	156,654.9	156,658.7	156,659.2	BIC Score	132,353.1	132353.6	132357.2
Weight	0.69	0.10	0.08	Weight	0.45	0.35	0.06
$\triangle$BIC	0	3.89	4.34	$\triangle$BIC	0	0.52	4.20
No. of Parameters	17	18	16	No. of Parameters	17	18	18

package. This study carries out a 10-fold CV using the Sobol sampling data. Similar to BIC model selection method, the data splitting employs a 50/50 split between training and testing data. The results of the LASSO FS method correspond to the complexity parameter obtained from CV (Table 10 and Table 11).

The selected features for energy consumption results conform with the Sobol sensitivity rankings (Table 10). However, there is a contradiction between LASSO results for the semi-detached archetype and the corresponding Sobol sensitivity rankings. The semi-detached LASSO results include the Roof_abs parameter and exclude the DHW parameter. The Sobol rankings indicate that the DHW parameter should be excluded prior to the omission of the Roof_abs parameter. While the LASSO results for the semi-detached archetype do not correspond with Sobol rankings, these results conform with the Morris sensitivity rankings. Also, the Sobol ranking of the Roof_abs and DHW parameters only differ by approximately 0.04% of the output variance, which negates the discrepancy between these results. Furthermore,

Table 10: Features Selected via LASSO Regression Based on the Energy Consumption (λ = penalization penalty at 1 std deviation from minimum error)

Bungalow		Detached		Semi-Detached		Top-Floor Apartment		Mid-Floor Apartment	
Sobol Rankings	LASSO Features	Sobol Rankings	LASSO Features	Sobol Rankings	LASSO Features	Sobol Rankings	LASSO Features	Sobol Rankings	LASSO Features
COP	✓	COP	✓	COP	✓	COP	✓	COP	✓
HSPT	✓	HSPT	✓	HSPT	✓	ACR	✓	ACR	✓
ACR	✓	Occ	✓	ACR	✓	HSPT	✓	Occ	✓
Occ	✓	ACR	✓	Occ	✓	Occ	✓	HSPT	✓
UGlaze	✓	UWall	✓	UGlaze	✓	URoof	✓	Equip	✓
Equip	✓	Equip	✓	Equip	✓	Equip	✓	UGlaze	✓
UWall	✓	UGlaze	✓	SHGC	✓	UGlaze	✓	SHGC	✓
SHGC	✓	SHGC	✓	UWall	✓	Orientation	✗	Orientation	✗
WWR	✓	WWR	✓	WWR	✓	SHGC	✓	UWall	✓
URoof	✓	URoof	✓	Ld	✓	DHW	✓	DHW	✓
Ld	✓	Ld	✓	URoof	✓	UWall	✓	WWR	✓
DHW	✓	Wall_abs	✓	Wall_abs	✓	Ld	✓	Ld	✓
Wall_abs	✓	DHW	✓	Roof_abs	✗	WWR	✗	Aux	✗
Roof_E	✓	Wall_E	✗	DHW	✓	Roof_E	✗	Wall_abs	✗
Roof_abs	✓	Aux	✗	Orientation	✗	Roof_abs	✗	Wall_E	✗
Wall_E	✗	Roof_abs	✗	Aux	✗	Aux	✗	HSBT	✗
Aux	✗	UFrame	✗	Roof_E	✗	Wall_abs	✗	UFrame	✗
UDoor	✗	Roof_E	✗	UFloor	✗	Wall_E	✗	UPart	✗
UFloor	✗	UFloor	✗	Wall_E	✗	HSBT	✗	Vt	✗
UFrame	✗	UDoor	✗	UFrame	✗	UFrame	✗	-	
Orientation	✗	Orientation	✗	UDoor	✗	UPart	✗	-	
UPart	✗	HSBT	✗	HSBT	✗	Vt	✗	-	
HSBT	✗	UPart	✗	UPart	✗	-		-	
Vt	✗	Vt	✗	Vt					
xmark	-		-						
λ = 34.2		λ = 33.5		λ = 28.8		λ= 17.3		λ = 15.7	

this discrepancy is not observed in the heating demand LASSO results for the semi-detached archetype that omit a similar set of non-influential parameters (Table 11). The discrepancies in Sobol energy consumption rankings and LASSO results could be simply due to a minor experimental error.

When observing the LASSO results for top and mid-floor apartments, the parameter set excludes the building orientation parameter from the list of influential parameters. However, the Sobol and Morris SA identify building orientation as an influential parameter. One potential reason for the exclusion could be due to the highly non-linear effect of this parameter on the model output. Other parameters such as the WWR also possess similar relationships with the model output but are not excluded from the top set of features determined by the LASSO method. Additionally, the LASSO method also determines the influential non-linear parameters when employing a large sample size [49]. This study uses 102,400 samples that represent a significant sample size. Hence, exclusion of the orientation parameter is not due to the sample size and requires further research.

Table 11: Features Selected via LASSO Regression Based on the Heating Demand (λ = penalization penalty at 1 std dev from minimum error)

Bungalow		Detached		Semi-Detached		Top-Floor Apartment		Mid-Floor Apartment	
Sobol Rankings	LASSO Features	Sobol Rankings	LASSO Features	Sobol Rankings	LASSO Features	Sobol Rankings	LASSO Features	Sobol Rankings	LASSO Features
HSPT	✓	HSPT	✓	HSPT	✓	ACR	✓	ACR	✓
ACR	✓	Occ	✓	ACR	✓	HSPT	✓	Occ	✓
Equip	✓	ACR	✓	Occ	✓	Occ	✓	HSPT	✓
Occ	✓	Equip	✓	Equip	✓	URoof	✓	Equip	✓
UGlaze	✓	UWall	✓	UGlaze	✓	Equip	✓	UGlaze	✓
UWall	✓	UGlaze	✓	UWall	✓	UGlaze	✓	SHGC	✓
SHGC	✓	SHGC	✓	SHGC	✓	SHGC	✓	Orientation	✓
WWR	✓	WWR	✓	WWR	✓	Orientation	✗	UWall	✗
URoof	✓	Ld	✓	Ld	✓	DHW	✓	DHW	✓
Ld	✓	URoof	✓	URoof	✓	UWall	✓	WWR	✓
DHW	✓	Wall_abs	✓	DHW	✓	Ld	✓	Ld	✓
Wall_abs	✓	DHW	✓	Orientation	✓	WWR	✗	Wall_abs	✗
Roof_abs	✓	Wall_E	✗	Wall_abs	✗	Roof_abs	✗	Wall_E	✗
Roof_E	✓	Roof_abs	✗	Wall_E	✓	Roof_E	✗	UFrame	✗
Wall_E	✗	UFrame	✗	Roof_abs	✗	Wall_abs	✗	UPart	✗
UDoor	✗	UFloor	✗	UFloor	✗	Wall_E	✗	HSBT	✗
UFloor	✗	Roof_E	✗	Roof_E	✗	HSBT	✗	Vt	✗
HSBT	✗	Orientation	✗	UFrame	✗	UFrame	✗	-	✗
Orientation	✗	UDoor	✗	UDoor	✗	UPart	✗	-	✗
UFrame	✗	HSBT	✗	UPart	✗	Vt	✗	-	
UPart	✗	UPart	✗	HSBT	✗	-	✗	-	
Vt	✗	Vt	✗	Vt	✗	-	✗	-	
$\lambda = 6.8$		$\lambda = 9.3$		$\lambda = 8.1$		$\lambda = 5.5$		$\lambda = 6.1$	

3.5.3. Evaluation of BIC and LASSO Regression Models

The evaluation of BIC and LASSO regression models uses RMSE and MPE evaluation metrics. These metrics compute the training and testing errors of both energy consumption and heating demand models for each archetype. The LASSO energy model is further evaluated using the following criteria:

1. Underfitting: Testing and training error is high;
2. Overfitting: Testing error is high and training error low;
3. Good fit: Testing error low, slightly higher than the training error and;
4. Unknown fit: Testing error low, training error high.

Bungalow Archetype: The BIC and LASSO energy models for the bungalow archetype are reasonably comparable in their prediction accuracy with only 0.12% difference in their respective MPEs (Table 12). The LASSO model offers a minor reduction in model complexity compared to the BIC model. Furthermore, the LASSO model retains and excludes parameters in a manner that correspond to the sensitivity rankings. Even though the LASSO model attains a marginal drop in prediction accuracy, this model provides a better description of the more influential parameters within the bungalow archetype. It should also be noted that the accuracy of these models does not necessarily represent the actual accuracy obtained through the actual use of these parameters. The LASSO energy model for the bungalow archetype can be considered a ‘good fit’ since both the training and testing errors are below 10% with the testing error being slightly higher than the training error. Therefore, the model is sufficiently generalized and is representative of the Irish housing stock.

The BIC and LASSO heating demand models are also similar in their predictive performance, although the difference in accuracy is slightly more prominent than that observed for the energy models (Table 13). The LASSO method further offers a reduction in model complexity and includes less parameters than the BIC model. The training and testing errors of BIC and LASSO models appear to have a ‘good fit’ with the data and hence, these models are generalizable for the Irish housing stock. Also, both BIC and LASSO models conform with the sensitivity rankings; the model selection merely depends on the desired accuracy. The motivation behind this study is to reduce model complexity to aid modelers, which leads to question whether the 0.95% decrease in predictive accuracy is worth the reduction in model complexity.

Table 12: Final Energy Consumption Models for each Archetype & their Associated RMSE & MPE

Bungalow		Detached		Semi-Detached		Top-Floor Apartment		Mid-Floor Apartment	
BIC	LASSO	BIC	LASSO	BIC	LASSO	BIC	LASSO	BIC	LASSO
COP	COP	COP	COP	COP	COP	COP	COP	COP	COP
HSPT	HSPT	HSPT	HSPT	HSPT	HSPT	ACR	ACR	ACR	ACR
ACR	ACR	Occ	Occ	ACR	ACR	HSPT	HSPT	Occ	Occ
Occ	Occ	ACR	ACR	Occ	Occ	Occ	Occ	HSPT	HSPT
UGlaze	UGlaze	UWall	UWall	UGlaze	UGlaze	URoof	URoof	Equip	Equip
Equip	Equip	Equip	Equip	Equip	Equip	Equip	Equip	UGlaze	UGlaze
UWall	UWall	UGlaze	UGlaze	SHGC	SHGC	UGlaze	UGlaze	SHGC	SHGC
SHGC	SHGC	SHGC	SHGC	UWall	UWall	Orientation	SHGC	Orientation	UWall
WWR	WWR	WWR	WWR	WWR	WWR	SHGC	DHW	UWall	DHW
URoof	URoof	URoof	URoof	Ld	Ld	DHW	UWall	DHW	WWR
Ld	Ld	Ld	Ld	URoof	URoof	UWall	Ld	WWR	Ld
DHW	DHW	Wall_abs	Wall_abs	Wall_abs	Wall_abs	Ld	-	Ld	-
Wall_abs	Wall_abs	DHW	DHW	Roof_abs	DHW	WWR	-	Aux	-
Roof_abs	Roof_E	Wall_E	-	DHW	-	Roof_E	-	Wall_abs	-
Wall_E	Roof_abs	Aux	-	Orientation	-	Roof_abs	-	-	-
Aux	-	Roof_abs	-	Aux	-	Aux	-	-	-
-	-	-	-	Wall_E	-	-	-	-	-
Training		Training		Training		Training		Training	
RMSE	RMSE	RMSE	RMSE	RMSE	RMSE	RMSE	RMSE	RMSE	RMSE
1298.2	1333.8	1703.4	1756.9	1412.7	1453.2	684.8	724.7	566.1	590.4
MPE	MPE	MPE	MPE	MPE	MPE	MPE	MPE	MPE	MPE
8.71%	8.74%	8.80%	8.82%	8.92%	8.92%	9.71%	9.94%	9.41%	9.51%
Testing		Testing		Testing		Testing		Testing	
RMSE	RMSE	RMSE	RMSE	RMSE	RMSE	RMSE	RMSE	RMSE	RMSE
1298.1	1338.6	1709.8	1742.5	1415.1	1444.1	682.63	729.8	568.5	590.9
MPE	MPE	MPE	MPE	MPE	MPE	MPE	MPE	MPE	MPE
8.73%	8.85%	8.82%	8.85%	8.93%	8.97%	9.81%	9.98%	9.47%	9.48%

Detached Archetype: The BIC and LASSO energy models for the detached archetype are very similar in their prediction accuracy with only 0.02% difference in their MPEs (Table 12). The LASSO model comprises three fewer parameters than the BIC where the extra three parameters, namely Wall_E, Aux and Roof_abs, are potentially difficult to define. The energy model conforms with the Sobol sensitivity rankings and represents a 'good fit' with the data. Also, the predictive accuracy of energy models are within the 10% range. Therefore, model selection again depends on the trade-off between model complexity and accuracy. The BIC and LASSO heating demand models also observe similar trends (Table 13). The LASSO heating demand model is less accurate than the BIC model with a similar reduction in model complexity. Both BIC and LASSO models render a 'good fit' with the data as per the relative training and testing MPEs.

Table 13: Final Heating Demand Models for each Archetype & their Associated RMSE & MPE

Bungalow		Detached		Semi-Detached		Top-Floor Apartment		Mid-Floor Apartment	
BIC	LASSO	BIC	LASSO	BIC	LASSO	BIC	LASSO	BIC	LASSO
HSPT	HSPT	HSPT	HSPT	HSPT	HSPT	ACR	ACR	ACR	ACR
ACR	ACR	Occ	Occ	ACR	ACR	HSPT	HSPT	Occ	Occ
Equip	Equip	ACR	ACR	Occ	Occ	Occ	Occ	HSPT	HSPT
Occ	Occ	Equip	Equip	Equip	Equip	URoof	URoof	Equip	Equip
UGlaze	UGlaze	UWall	UWall	UGlaze	UGlaze	Equip	Equip	UGlaze	UGlaze
UWall	UWall	UGlaze	UGlaze	UWall	UWall	UGlaze	UGlaze	SHGC	SHGC
SHGC	SHGC	SHGC	SHGC	SHGC	SHGC	SHGC	SHGC	Orientation	UWall
WWR	WWR	WWR	WWR	WWR	WWR	Orientation	DHW	UWall	DHW
URoof	URoof	Ld	Ld	Ld	Ld	DHW	UWall	DHW	WWR
Ld	Ld	URoof	URoof	URoof	URoof	UWall	Ld	WWR	Ld
DHW	DHW	Wall_abs	Wall_abs	DHW	DHW	Ld	-	Ld	-
Wall_abs	Wall_abs	DHW	DHW	Orientation	Orientation	WWR	-	Wall_abs	-
Roof_abs	Roof_abs	Wall_E	-	Wall_abs	-	Roof_abs	-	Wall_E	-
Roof_E	Roof_E	Roof_abs	-	Wall_E	-	Roof_E	-	-	-
Wall_E	-	-	-	Roof_abs	-	Wall_abs	-	-	-
UDoor	-	-	-	UFloor	-	Wall_E	-	-	-
-	-	-	-	Roof_E	-	-	-	-	-
-	-	-	-	UDoor	-	-	-	-	-
Training		Training		Training		Training		Training	
RMSE	RMSE	RMSE	RMSE	RMSE	RMSE	RMSE	RMSE	RMSE	RMSE
347.2	469.7	571.5	669.6	459.5	553.2	290.9	407.3	298.6	590.4
MPE	MPE	MPE	MPE	MPE	MPE	MPE	MPE	MPE	MPE
3.13%	4.06%	4.69%	5.06%	4.22%	4.73%	6.26%	8.22%	8.88%	8.84%
Testing		Testing		Testing		Testing		Testing	
RMSE	RMSE	RMSE	RMSE	RMSE	RMSE	RMSE	RMSE	RMSE	RMSE
353.0	469.8	566.6	670.9	458.8	550.1	293.3	403.8	294.1	590.9
MPE	MPE	MPE	MPE	MPE	MPE	MPE	MPE	MPE	MPE
3.15%	4.10%	4.73%	5.13%	4.22%	4.78%	6.33%	8.13%	8.75%	8.95%

Semi-detached Archetype: Similar to previous archetypes, the predictive MPE of the LASSO energy model is higher than the BIC model (Table 12). Each model also represents a good fit to the data. However, the BIC model does not entirely agree with the sensitivity rankings. The LASSO feature selection method significantly reduces the model complexity compared to the BIC model with four less parameters. One of the excluded parameters in the LASSO method is the building orientation parameter. The orientation parameter does not substantially reduce the modeling time for the modeler. However, the exclusion of parameters such as DHW, Aux and Wall_E parameters potentially reduced the required modeling time. Regardless of the energy model selected in Table 33, both models lie within the 10% accuracy range.

The BIC heating demand model represents an excellent fit to the data with identical training and testing MPEs (Table 13). Despite a slight difference in the training and testing MPEs of the LASSO model, the model still represents a good fit to the data. One notable difference is the significant reduction in model complexity of the LASSO feature selection method. The six additional parameters in the BIC model only increase the predictive accuracy by 0.56% compared to the LASSO model. The negligible increase in predictive accuracy indicates that the additional parameters in the BIC model are non-influential.

Top-Floor Apartment: The energy models for the top-floor apartment exhibit similar characteristics to previous archetype models except the MPEs are slightly higher and just about falls under the 10% range (Table 12). Both BIC and LASSO energy models still represent a good fit to the data with moderate training and testing errors. The difference in the MPE between the BIC and LASSO heating demand models is significantly larger than that experienced in previous archetypes (Table 13). The reduced complexity of the LASSO heating demand model comes at the cost of an additional 1.8% decrease in the predictive accuracy. It can also be seen that the training error for the LASSO heating demand model is higher than the testing error thus characterising the fit as ‘unknown’. This could be a result of how data set is split (i.e. 50/50% split). The training set could have had perhaps many ‘hard’ cases to learn while the testing set had mostly ‘easy’ cases to predict.

Mid-Floor Apartment: Similar to the top-floor apartment, the LASSO energy model represents an ‘unknown fit’ for the aforementioned reasons (Table 12). The predictive accuracy of the BIC and LASSO energy models are nearly identical with only a 0.01% difference. Although, the LASSO model ,again, provides a reduced level of model complexity. These models just fall under the 10% accuracy threshold similar to the top-floor apartment. The heating demand models observe similar trends and observations as the previous archetypes (Table 13).

3.6. Data Output & Exchange

This process combines the results of SA and FS techniques to formulate minimum datasets of influential parameters (Table 14 and Table 15). The results obtained from SA and FS methods correspond/agree with one another. However, these results do not necessarily define a minimum data set that is optimal. Therefore, this study introduce further assumptions to develop the minimum data set for each archetype.

Bungalow Archetype: When assessing the overall results for energy consumption parameters, a few notable points can be made. The BIC energy model does not completely agree with the Morris or Sobol SA rankings. However, the LASSO energy model agrees with both Morris and Sobol SA results. The LASSO FS method identifies one less parameter than the BIC method with an additional 0.12% increase in MPE. The additional parameter in question is the auxiliary energy, which relates to a 0.045% impact on output variance according to the Sobol sensitivity results. The addition of this parameter could 'potentially' increase the accuracy of the bungalow archetype model through informed assumptions. However, the time required to accurately define this parameter is not deemed to be worthwhile, therefore, the LASSO energy consumption results are chosen as the final minimum data set for the bungalow archetype. Similar inferences could be used to formulate the heating demand minimum data set. As this study focuses more on models with reduced complexity, the LASSO model is chosen as the final subset of parameters to describe the heating demand.

Table 14: Guideline minimum set of accurately-defined input data for Energy Consumption (kWh)

Rank	Bungalow	Detached	Semi-Detached	Top-Flat	Mid-Flat
1	COP	COP	COP	COP	COP
2	HSPT	HSPT	HSPT	ACR	ACR
3	ACR	Occ	ACR	HSPT	Occ
4	Occ	ACR	Occ	Occ	HSPT
5	UGlaze	UWall	UGlaze	URoof	Equip
6	Equip	Equip	Equip	Equip	UGlaze
7	UWall	UGlaze	SHGC	UGlaze	SHGC
8	SHGC	SHGC	UWall	SHGC	UWall
9	WWR	WWR	WWR	Orientation	DHW
10	URoof	URoof	Ld	DHW	WWR
11	Ld	Ld	URoof	UWall	Ld
12	DHW	Wall_abs	Wall_abs	Ld	
13	Wall_abs	DHW	DHW		
14	Roof_E				
15	Roof_abs				

Detached Archetype: The subset of energy parameters produced by the BIC and LASSO FS methods for the detached archetype agree with the sensitivity rankings. The LASSO energy model provides a smaller subset of influential parameters compared to the BIC model with a 0.03% reduction in accuracy. Wall_E, Aux and Roof_abs are the three additional parameters omitted from the LASSO energy model which account for 0.049%, 0.048% and 0.037% of the output variance respectively. Given the marginal impact these parameters have on model accuracy, it could be concluded that the smaller subset of parameters in the LASSO energy model provides similar levels of accuracy as the BIC model. Henceforth, the LASSO model is chosen to define the final minimum data set. The BIC and LASSO heating demand models reflect identical scenarios as these models have similar levels of predictive accuracy and agree with the sensitivity results. Therefore, the LASSO model is chosen as the final subset of parameters to describe the heating demand.

Semi-detached Archetype: Overall, the FS results for the semi-detached archetype do not conform very well with the sensitivity results when compared to the bungalow and detached archetypes. Both the BIC and LASSO energy models disagreed with the sensitivity rankings, although, the LASSO results did conform with Morris rankings. Similar to previous archetypes, the minor loss in accuracy of the LASSO energy model is insignificant given the considerably large reduction in model complexity. Therefore, the LASSO model is chosen to define the set of influential parameters that formulate the minimum dataset. With regards to the heating demand, the LASSO model performs remarkably given the significant reduction in the number of parameters for a slight loss in the predictive accuracy. The BIC heating demand model fits the data slightly better than the LASSO model but at the expense of six additional parameters. The potential time required to obtain the necessary data for these parameters does not add sufficient information. Therefore, the LASSO parameter subset is the final minimum dataset for the semi-detached archetype.

Top and Mid-floor Apartments: The energy models for the top-floor apartment possess similar characteristics as the previous archetypes whereby BIC offers better accuracy at the cost of model complexity. The LASSO energy model might appear to be more favourable due to the smaller subset of influential parameters. However, the LASSO model excludes the building orientation parameter which has a moderate effect on the output variance (i.e. 1.5%). The LASSO model further includes parameters that are ranked lower in sensitivity rankings than the orientation parameter. A potential reason for excluding the orientation parameter might be due to the high non-linearity of this parameter. Therefore, given that the LASSO energy model agrees with the sensitivity rankings apart for the orientation parameter, the selected subset includes parameters identified by the LASSO method along with the orientation parameter. The corresponding BIC model involves a high degree of model complexity. Therefore, the minimum dataset consist of the parameters of the LASSO energy model along with the orientation parameter. The heating demand models experience a similar issue except there was a higher loss in the predictive accuracy between BIC and LASSO models. Including the orientation parameter could aid in bridging the gap in predictive accuracy. Therefore, the LASSO parameter subset is chosen with the added inclusion of the orientation parameter to represent the heating demand. The mid-floor apartment energy consumption and heating demand results are very similar except the relative model complexity is not as significant due to fewer overall parameters of the mid-floor apartment (adiabatic roof assumption). Therefore, the LASSO model defines the minimum dataset for mid-floor apartments along with the inclusion of the orientation parameter.

4. Discussions

While the two feature selection methods produced similar results, the LASSO models were consistently less complex than the BIC models. However, the LASSO models were less accurate but the loss in accuracy was marginal (differences in the predictive accuracy range between 0.01% and 1%) , therefore, leading to the subset of parameters produced by the LASSO method to be the final minimum data set of parameters. While the LASSO models were less complex and also agreed with the sensitivity rankings better than the BIC approach, these models experience an issue detecting the influence of the building orientation parameter within the apartment archetypes. Building orientation was considered an influential parameter according to the sensitivity rankings. We proposed that this issue

Table 15: Guideline minimum set of accurately-defined input data for Heating Demand (kWh)

Rank	Bungalow	Detached	Semi-Detached	Top-Flat	Mid-Flat
1	HSPT	HSPT	HSPT	ACR	ACR
2	ACR	Occ	ACR	HSPT	Occ
3	Equip	ACR	Occ	Occ	HSPT
4	Occ	Equip	Equip	URoof	Equip
5	UGlaze	UWall	UGlaze	Equip	UGlaze
6	UWall	UGlaze	UWall	UGlaze	SHGC
7	SHGC	SHGC	SHGC	SHGC	Orientation
8	WWR	WWR	WWR	Orientation	UWall
9	URoof	Ld	Ld	DHW	DHW
10	Ld	URoof	URoof	UWall	WWR
11	DHW	Wall_abs	DHW	Ld	Ld
12	Wall_abs	DHW	Orientation		
13	Roof_abs				
14	Roof_E				

arose due to the non-linear and/or non-monotonic behaviour of the orientation parameter, however, further research is required to determine the root cause of this problem. With the sensitivity results clearly indicating the importance, the building orientation parameter was simply added to the subset of parameters already chosen by the LASSO method to counteract this issue.

While local sensitivity analysis can provide the top cluster of important parameters, this method fails to segregate the influential and non-influential parameters. This can be problematic since the rankings are deemed inaccurate. Global sensitivity analysis techniques are advocated and are becoming the common choice of analysis in building related studies. When trying to define a minimum dataset of influential parameters, the BIC approach is more beneficial due to the greater penalization on model complexity.

5. Conclusions and Future Work

As new design ideas push the envelope of building performance, the performance evaluation needs to be justified by building energy performance simulation. It has become crucial to investigate the accuracy, validity and relevance of developed models. These models involve several thousand input parameters; each parameter has different influence on the building energy consumption. Therefore, the need for identifying influential parameters is quite evident to facilitate the building energy modeling process. The performance gap commonly experienced between predicted and actual energy performance is a topic of continuing research. One such factor causing this performance gap is modeling illiteracy. Studies often define a minimum dataset including the most influential building energy performance parameters to aid modellers in the model development process. However, while somewhat successful, the previous research does not provide a clear methodology to accurately identify and validate the influential or non-influential parameters in relation to the predicted energy consumption.

This study proposes an integrated sequential methodology to identify the most influential parameters through the use of global sensitivity analysis and feature selection techniques. The methodology demonstrates an integrated use of Bayesian Information Criteria and an embedded feature selection algorithm called the LASSO method. The results of these feature selection methods are tested and validated through the use of two global sensitivity analysis techniques, namely Morris and Sobol sensitivity analysis. Each set of influential parameters determined by feature selection techniques have predicative accuracy that are within 10% of the actual results.

While the guideline minimum data set produced in this study is only applicable to Irish dwellings, the proposed methodology can be applied to various types of residential building models. jEPlus and EnergyPlus are highly versatile simulation tools, which can assess various building parameters and while some parameters might not be entirely applicable there are several approaches to overcome these issues similar to procedures taken in this work.

The guideline minimum datasets can be advantageous during numerous stages of building energy performance simulation over a building's life-cycle. The time spent collecting and defining parameter input data, along with the required modeling time, can be significantly reduced by identifying and accurately modeling only the most influential input parameters. In the late design and/or commissioning stages of a building's life-cycle, when a high degree of modeling accuracy is required, the minimum data set can be referenced in order to ensure that the most influential input parameters are accurately defined. The devised methodology can be applied to other building stocks. The formulated datasets represent only influential features and hence, could be used by urban planners and energy policymakers to estimate energy retrofit investment costs, emission reductions and energy savings.

Future work could investigate the effect of the building orientation parameter on the LASSO regression feature selection method. Although issues were experienced with parameters that exhibited a highly non-linear and/or non-monotonic relationship with the model output, determining these issues is an area for potential future work. While the orientation parameter is categorized as an influential parameter in the sensitivity analysis, the feature selection method excludes this parameter from the subset of influential parameters. With the above issue experienced in the LASSO regression method, an additional path would be to investigate other feature selection algorithms. One potential algorithm that would be worth investigating is Principal component analysis, which generates linear combinations of original features and removes the effect of correlated variables on the identified list of influential variables. The first principal component explains the most variance in a dataset, and the second component explains the second-most variance, and this continues for each principal component. A more complete analysis of the Irish building sector could be obtained by applying the methodology used in this study to the Irish commercial building stock. This could potentially result in a data set that is capable of accurately defining both commercial and residential Irish buildings.

6. Acknowledgements

This publication has emanated from research supported in part by a research grant from Science Foundation Ireland (SFI) under the SFI Strategic Partnership Programme Grant number SFI/15/SPP/E3125. The opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the Science Foundation Ireland.

The author would also like to acknowledge, IBPSA Project 1, an international project conducted under the umbrella of the International Building Performance Simulation Association (IBPSA). Project 1 will develop and demonstrate a BIM/GIS and Modelica Framework for building and community energy system design and operation.

References

[1] T. Abergel, C. Delmastro, Tracking Buildings 2020, Tech. rep., International Energy Agency, Paris (2020).
URL `https://www.iea.org/reports/tracking-buildings-2020`

[2] IEA, Sustainable Development Scenario – World Energy Model, Tech. rep., IEA (November 2019).
URL `https://www.iea.org/reports/world-energy-model/sustainable-development-scenario`

[3] European Union, EU energy in figures - Statistical Pocketbook, Tech. rep., European Union (2020). `doi:10.2833/75283`.

[4] M. H. Shamsi, U. Ali, J. O'Donnell, A generalization approach for reduced order modelling of commercial buildings, Journal of Building Performance Simulation 12 (6) (2019) 729–744. `doi:10.1080/19401493.2019.1641554`.
URL `https://www.tandfonline.com/doi/abs/10.1080/19401493.2019.1641554`

[5] F. Salata, V. Ciancio, J. Dell'Olmo, I. Golasi, O. Palusci, M. Coppi, Effects of local conditions on the multi-variable and multi-objective energy optimization of residential buildings using genetic algorithms, Applied Energy 260 (2020) 114289. `doi:10.1016/j.apenergy.2019.114289`.

[6] G. Yang, Z. Li, G. Augenbroe, Development of prototypical buildings for urban scale building energy modeling: A reduced order energy model approach, Science and Technology for the Built Environment 24 (1) (2018) 33–42. `doi:10.1080/23744731.2017.1328943`.

[7] D. Coakley, P. Raftery, M. Keane, A review of methods to match building energy simulation models to measured data, Renewable and Sustainable Energy Reviews 37 (2014) 123–141. `arXiv:j.rser.2014.05.007`, `doi:10.1016/j.rser.2014.05.007`.
URL `https://www.sciencedirect.com/science/article/pii/S1364032114003232{#}!`

[8] J. H. Choi, Investigation of the correlation of building energy use intensity estimated by six building performance simulation tools, Energy and Buildings 147 (2017) 14–26. `doi:10.1016/j.enbuild.2017.04.078`.

[9] P. de Wilde, R. Jones, The building energy performance gap: Up close and personal, in: Proceedings of the CIBSE ASHRAE Technical Symposium: Moving to a new world of building systems performance, Dublin, Ireland, 2014, pp. 1–8.
URL `http://hdl.handle.net/10026.1/4324`

[10] C. van Dronkelaar, M. Dowson, C. Spataru, D. Mumovic, A Review of the Regulatory Energy Performance Gap and Its Underlying Causes in Non-domestic Buildings, Frontiers in Mechanical Engineering 1 (2016) 17. `doi:10.3389/fmech.2015.00017`.
URL `http://journal.frontiersin.org/Article/10.3389/fmech.2015.00017/abstract`

[11] P. Shiel, S. Tarantino, M. Fischer, Parametric analysis of design stage building energy performance simulation models, Energy and Buildings 172 (2018) 78–93. `doi:10.1016/j.enbuild.2018.04.045`.

[12] S. Imam, D. A. Coley, I. Walker, The building performance gap: Are modellers literate?, Building Services Engineering Research and Technology 38 (3) (2017) 351–375. `doi:10.1177/0143624416684641`.
URL `http://journals.sagepub.com/doi/10.1177/0143624416684641`

[13] J. W. Hand, D. B. Crawley, M. Donn, L. K. Lawrie, Improving non-geometric data available to simulation programs, Building and Environment 43 (4) (2008) 674–685. `doi:10.1016/j.buildenv.2006.06.022`.

[14] D. Garcia Sanchez, B. Lacarrière, M. Musy, B. Bourges, Application of sensitivity analysis in building energy simulations: Combining first- and second-order elementary effects methods, Energy and Buildings 68 (PART C) (2014) 741–750. `arXiv:1203.3055`, `doi:10.1016/j.enbuild.2012.08.048`.

[15] R. Arababadi, H. Naganathan, K. Parrish, W. K. Chong, Determining the Feasibility of Statistical Techniques to Identify the Most Important Input Parameters of Building Energy Models, in: Procedia Engineering, Vol. 118, Elsevier Ltd, 2015, pp. 1258–1265. `doi:10.1016/j.proeng.2015.08.478`.

[16] T. Wei, A review of sensitivity analysis methods in building energy analysis (apr 2013). `doi:10.1016/j.rser.2012.12.014`.

[17] K. Menberg, Y. Heo, R. Choudhary, Sensitivity analysis methods for building energy models: Comparing computational costs and extractable information, Energy and Buildings 133 (2016) 433–445. `doi:10.1016/j.enbuild.2016.10.005`.

[18] M. H. Kristensen, S. Petersen, Choosing the appropriate sensitivity analysis method for building energy model-based investigations, Energy and Buildings 130 (2016) 166–176. `doi:10.1016/j.enbuild.2016.08.038`.

[19] J. Egan, D. Finn, P. H. Deogene Soares, V. A. Rocha Baumann, R. Aghamolaei, P. Beagon, O. Neu, F. Pallonetto, J. O'Donnell, Definition of a useful minimal-set of accurately-specified input data for Building Energy Performance Simulation, Energy and Buildings 165 (2018) 172–183. `doi:10.1016/j.enbuild.2018.01.012`.

[20] A. Saltelli, S. Tarantola, F. Campolongo, M. Ratto, Sensitivity Analysis in Practice, John Wiley & Sons, Ltd, Chichester, UK, 2002. `doi:10.1002/0470870958`.
URL `http://doi.wiley.com/10.1002/0470870958`

[21] P. Heiselberg, H. Brohus, A. Hesselholt, H. Rasmussen, E. Seinre, S. Thomas, Application of sensitivity analysis in design of sustainable buildings, Renewable Energy 34 (9) (2009) 2030–2036. `doi:10.1016/j.renene.2009.02.016`.

[22] K. Menberg, Y. Heo, R. Choudhary, Sensitivity analysis methods for building energy models: Comparing computational costs and extractable information, Energy and Buildings 133 (2016) 433–445. `doi:10.1016/j.enbuild.2016.10.005`.

[23] P. De Wilde, W. Tian, Predicting the performance of an office under climate change: A study of metrics, sensitivity and zonal resolution, Energy and Buildings 42 (10) (2010) 1674–1684. `doi:10.1016/j.enbuild.2010.04.011`.

[24] L. Wang, R. Kubichek, X. Zhou, Adaptive learning based data-driven models for predicting hourly building energy use, Energy and Buildings 159 (2018) 454–461. `doi:10.1016/j.enbuild.2017.10.054`.

[25] J. A. Carta, P. Cabrera, J. M. Matías, F. Castellano, Comparison of feature selection methods using ANNs in MCP-wind speed methods. A case study, Applied Energy 158 (2015) 490–507. `doi:10.1016/j.apenergy.2015.08.102`.

[26] G. Chandrashekar, F. Sahin, A survey on feature selection methods, Computers and Electrical Engineering 40 (1) (2014) 16–28. `doi:10.1016/j.compeleceng.2013.11.024`.

[27] U. Ali, M. H. Shamsi, M. Bohacek, C. Hoare, K. Purcell, E. Mangina, J. O'Donnell, A data-driven approach to optimize urban scale energy retrofit decisions for residential buildings, Applied Energy 267 (2020) 114861. `doi:10.1016/j.apenergy.2020.114861`.

[28] TABULA-typology approach for building stock energy assessment, EPISCOPE, `http://episcope.eu`, [Online; accessed 2-Mar-2020].

[29] jEPlus User Manual.
URL `http://www.jeplus.org/wiki/doku.php?id=docs:manual{_}1{_}7`

[30] T. A. Mara, S. Tarantola, Application of global sensitivity analysis of model output to building thermal

simulations, Building Simulation 1 (4) (2008) 290–302. doi:10.1007/s12273-008-8129-5.
URL https://link.springer.com/article/10.1007/s12273-008-8129-5

[31] C. Tong, F. Graziani, A practical global sensitivity analysis methodology for multi-physics applications, in: Lecture Notes in Computational Science and Engineering, Vol. 62 LNCSE, Springer, Berlin, Heidelberg, 2008, pp. 277–299. doi:10.1007/978-3-540-77362-7_12.
URL https://link.springer.com/chapter/10.1007/978-3-540-77362-7{_}12

[32] R. A. Cropp, R. D. Braddock, The New Morris Method: An efficient second-order screening method, Reliability Engineering and System Safety 78 (1) (2002) 77–83. doi:10.1016/S0951-8320(02)00109-6.

[33] I. M. Sobol, Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates, Mathematics and Computers in Simulation 55 (1-3) (2001) 271–280. doi:10.1016/S0378-4754(00)00270-6.

[34] Y. Zhang, S. Ray, W. Gurj, On the consistency of feature selection with Lasso for non-linear targets, in: 33rd International Conference on Machine Learning, ICML 2016, Vol. 1, 2016, pp. 322–330.

[35] R. Tibshirani, Regression shrinkage and selection via the lasso: A retrospective, Journal of the Royal Statistical Society. Series B: Statistical Methodology 73 (3) (2011) 273–282. doi:10.1111/j.1467-9868.2011.00771.x.

[36] C. Cui, D. Wang, High dimensional data regression using Lasso model and neural networks with random weights, Information Sciences 372 (2016) 505–517. doi:10.1016/j.ins.2016.08.060.

[37] S. Wang, B. Ji, J. Zhao, W. Liu, T. Xu, Predicting ship fuel consumption based on LASSO regression, Transportation Research Part D: Transport and Environment 65 (2018) 817–824. doi:10.1016/j.trd.2017.09.014.

[38] O. Neu, S. Oxizidis, D. Flynn, D. Finn, Utilising Time of Use Surveys to Predict Domestic Hot Water Consumption and Heat Demand Profiles of Residential Building Stocks, British Journal of Environment and Climate Change 6 (2) (2016) 77–89. doi:10.9734/bjecc/2016/18188.
URL http://hdl.handle.net/10197/8007

[39] P. Washan, D. Ross, Prepared for: Department of Housing, Planning and Local Government AECOM Quality information, Tech. rep., Department of Housing, Planning and Local Government (2019).

[40] A. A. Famuyibo, A. Duffy, P. Strachan, Developing archetypes for domestic dwellings - An Irish case study, Energy and Buildings 50 (2012) 150–157. doi:10.1016/j.enbuild.2012.03.033.

[41] Census 2016 Profile 4 – Households and Families - CSO - Central Statistics Office.
URL https://www.cso.ie/en/csolatestnews/presspages/2017/cp4hf/

[42] SIMLAB and other software — EU Science Hub.
URL https://ec.europa.eu/jrc/en/samo/simlab

[43] F. Campolongo, J. Cariboni, Sensitivity analysis: how to detect important factors in large models, Tech. rep., European Union (2007).
URL http://www.jrc.cec.eu.int

[44] A. T. Nguyen, S. Reiter, A performance comparison of sensitivity analysis methods for building energy models, Building Simulation 8 (6) (2015) 651–664. doi:10.1007/s12273-015-0245-4.
URL https://link.springer.com/article/10.1007/s12273-015-0245-4

[45] I. Korolija, L. Marjanovic-Halburd, Y. Zhang, V. I. Hanby, Influence of building parameters and HVAC systems coupling on building energy performance, Energy and Buildings 43 (6) (2011) 1247–1253. doi:10.1016/j.enbuild.2011.01.003.

[46] BARTON, K, MuMIn : multi-model inference. R package version 1. 0. 0. (2009).
URL http://ci.nii.ac.jp/naid/10030918982/en/

[47] K. P. Burnham, D. R. Anderson, Model Selection and Multimodel Inference, Springer New York, 2004. doi:10.1007/b97636.

[48] J. Friedman, T. Hastie, R. Tibshirani, Regularization Paths for Generalized Linear Models via Coordinate Descent, Tech. rep., Stanford University (2009).

[49] Y. Zhang, S. Ray, W. Gurj, On the consistency of feature selection with Lasso for non-linear targets, in: 33rd International Conference on Machine Learning, ICML 2016, Vol. 1, 2016, pp. 322–330.

Appendix A. Description of the Irish residential archetypes

Table Appendix A.1: Physical Building Parameters describing the different archetypes.

Building Parameters	Bungalow	Detached	Semi-detached	Top/Mid-floor Apartment	Units
Floor Area	85.91	130.81	105.47	45.8	m^2
No. of Zones	8	13	10	5	-
Glazed Area	18.6	31.84	21.5	8.65	m^2
Number of Windows	9	9	9	2	-
External Door Area	4.01	3.98	2.19	0	m^2
No. of Doors	2	2	2	2	-
External Wall Area	78.19	147.18	139.43	63.35	m^2

Appendix B. Morris Sensitivity Results

Morris SA Results for Detached Archetype: The set of influential parameters is similar to the one obtained for the Bungalow archetype when analyzing the building energy consumption (Table Appendix B.1). Similar to the bungalow archetype, the seasonal COP of the heating system is the most influential parameter in the detached model. Factors such as UGlaze, UWall and Wall_{abs} experience a change in rankings due to the reduction in average building WWR. The glazed portion U-value (UGlaze) was previously the fifth most influential parameter in the bungalow archetype and dropped to the seventh place for the detached house. Parameters such as external wall U-value and wall solar absorption experienced a rise in their rankings, which is strongly related to the rise in wall percentage and fall in glazed percentage.

The analysis of heating demand SA results indicate that both WWR and Wall_{abs} are situated in the 'almost monotonic' region. As mentioned earlier, WWR lies within the 'almost monotonic' region which suggests the presence of interaction effects. The wall solar absorption parameter has moved from the 'monotonic region' in the bungalow archetype to the almost monotonic region for the detached dwelling when the percentage of external wall has increased. The influence of building orientation on the model outcome is non-linear and/or non-monotonic which again indicates the presence of interaction effects. This would be expected given the orientation is related to all solar effected parameters. The ranking of parameters that influence the electrical load are the same as the bungalow archetype. The equipment density is still significantly influential than lighting density and auxiliary energy. The proportion of significance each parameter has on the electrical load remains consistent throughout all five dwelling archetypes since their influence is entirely dependent on the floor area. However, their importance relative to energy or heating load parameters may vary.

Morris SA Results for Semi-detached Archetype: The Morris sensitivity rankings for the semi-detached archetype are similar to the detached archetype with a few exceptions

(Table Appendix B.1). External wall U-value, orientation and roof U-value experience a change in their rankings respectively. The change in the rank of these parameters is a direct result of the adiabatic wall assumption for this building archetype. With the external wall area reduced due to the assumption of an adiabatic wall, the archetype observed a decrease in the influence of the UWall parameter. When comparing the rankings of the UWall parameter for detached and semi-detached dwellings, the ranking dropped from fourth to seventh place in the energy consumption sensitivity. The effect of the adiabatic wall is also seen on the ranking of the orientation parameter, which experienced a rise from 21st for the detached dwelling to 13th for the semi-detached dwelling. The adiabatic wall results in an extreme loss or gain depending on the orientation.

Morris SA Results for Top and Mid-floor Apartments: Comparing the top and mid-floor apartments to previous archetypes, the U-value of the roof is found to be more sensitive for the top-floor apartment (Table Appendix B.2). This increase in sensitivity could be partly due to the flat roof design of the top-floor apartment compared to the pitched roof scheme of previous dwellings. The loft zone should act as a thermal 'buffer' in the pitched roof design thus minimizing the overall effect of the roof U-value. Both top and mid-floor apartments experience similar physical effects in contrast to other archetypes. ACR is the second most influential parameter for energy consumption in both archetypes due to the decrease in sensitivity of the HSPT parameter. For previous archetypes, this parameter is consistently the most influential parameter behind COP. However, HSPT dropped to the fourth place in energy sensitivity rankings for both top and mid-floor apartments. This would be expected since maintaining a higher heating set-point temperature for voluminous archetypes such as the bungalow would be far more energy intensive. Furthermore, the building orientation is far more significant for top and mid-floor apartments compared to other dwellings. These apartments usually have a single glazed façade compared to the front and back glazed facades for other dwellings.

Appendix C. Sobol Sensitivity Indices

Sobol Results for the Detached Archetype: Similar to the bungalow archetype, the effect of the building orientation parameter is doubled when the interaction terms are taken into account (Table Appendix C.1). Although, the influence of the orientation parameter on energy consumption and heating demand is insignificant. Examining the Sobol indices for the detached dwelling, a noticeable decrease in the influence of the heating system COP is observed compared to the bungalow archetype. By analysing the energy breakdown, it is found that the heating demand accounted for 65% of the total load while the electrical demand accounted for 35%. The bungalow archetype has a 73/27% split on the total load. The 8% reduction in heating requirements decreases the overall effect of the heating system COP.

The electrical load has a higher effect on the detached dwelling as compared to the bungalow. Therefore, the equipment and lighting densities cover a greater fraction of the output variance for this dwelling type. The fraction that both these parameters exhibit has doubled. However, while the increased electrical load is an influential factor, the area of the building must also be considered. The bungalow archetype has a floor area of 85.91 m^2 while the detached dwelling has an area of 130.81 m^2. Since the magnitude of these parameters are dependent on floor space (i.e. W/m^2), the increased floor area is perhaps the biggest contributor to their increased sensitivity. Furthermore, the relative ranking of equipment and lighting density remains consistent throughout all dwellings due to their value being dependent on floor area.

With the reduction in heating load requirements, the influence of the UWall parameter increases. This confirms that the reduction in overall building glazed surfaces results in the UWall parameter being more dominant in the energy consumption outcome. As clearly seen in heating load rankings, the UWall parameter has doubles in sensitivity compared to the bungalow archetype growing from 2.1% to 4.4%. Similar to the bungalow archetype, the WWR and building orientation parameters for the detached dwelling still highly depend on its interaction effects with other parameters for both energy consumption and heating load.

Sobol Results for the Semi-detached Archetype: The Sobol SA results follow similar trends as the detached archetype (Table Appendix C.2). The COP of the heating system increases with respect to the growth in heating demand and corresponding reduction in electrical load. Conversely, the equipment density and auxiliary energy output variance experience a slight decrease. While the lighting density parameter decreased for both heating and electrical loads, this parameter experienced an increase for energy consumption. This could be merely due to issues with model convergence.

The assumption of the adiabatic wall has a similar influence on the Sobol SA results as for the Morris SA results. The influence of external wall U-value on the output variance is nearly halved for both energy consumption and heating load. This reduced sensitivity would be expected since an entire building façade that constituted a large proportion of the external wall is now no longer susceptible to variations in the UWall parameter. Furthermore, the total order sensitivities of wall solar absorption and emissivity have decreased by 42% and 46% from their detached sensitivity value.

The adiabatic wall also influences the building orientation parameter for similar reasons as described earlier. The building orientation output variance has increased by a factor of five for both energy consumption and heating demand when compared to the detached dwelling. While the building orientation effect has increased dramatically, the parameter still only accounts for approximately 0.06% and 0.2% of the output variance for energy consumption and heating demand. The sensitivity index of the roof U-value declined by approximately 33% when compared to the detached archetype. The reduction is largely due to the smaller fraction of the roof being exposed to the external environment. The roof of the detached dwelling has an external surface area of 115.7 m^2 compared to the roof of semi-detached dwelling that has an area of 81.8 m^2.

Sobol Results for Top & Mid-Floor Apartments: The heating set-point temperature of the top-floor apartment accounts for far less of the output variance than previous dwellings (Table Appendix C.3). The influence of this parameter on energy consumption and heating load is nearly one-third of previous archetypes. The bungalow, detached and semi-detached archetypes are far more voluminous than the top-floor apartment. Therefore, the heating set-point would be expected to reduce for the top-floor apartment owing to the reduced volume (approximately one-third of larger archetypes). The U-value of the roof for the top-floor apartment is substantially more sensitive to the output variance of the energy consumption and heating load. The effect on the output variance is greater than five times the sensitivity index of previous dwellings. Similarly, the building orientation accounts for a larger amount of the output variance for the top-floor apartment compared to the previous dwellings.

The sensitivity of the WWR parameter has approximately halved when compared to the detached and semi-detached archetypes. If the top-floor apartment is experiencing marginal solar gains through the glazed windows, then the WWR results in small amounts of variation in both the energy consumption and heating demand. Therefore, the effect of WWR is highly dependent on the building orientation. This is quite apparent when comparing the higher order effects of the WWR between the semi-detached dwelling and the top-floor apartment. The WWR sensitivity index doubles when including the higher order effects for the semi-detached dwelling. On the other hand, the WWR sensitivity index increases by a factor of 16 when including the higher order effects for the top-floor apartment.

The area of the top-floor apartment is 45.8 m^2, which is more than half the area of the semi-detached archetype (105.47 m^2). With this large reduction in area, the equipment density experiences a decrease in the sensitivity index. The lighting density, however, experiences an increase in the sensitivity index. The increase in the sensitivity of lighting density could be due to two possible reasons. Firstly, the increased sensitivity could be a result of convergence issues of small sensitivity indices. Secondly, this increase could be due to the higher modal value set in the top-floor apartment compared to previous archetypes, thus, resulting in a greater impact on energy consumption than usual. The second point appears to be more likely since the sensitivity index of the electrical load for the top-floor apartment is greater than that experienced in the semi-detached archetype.

The mid-floor apartment experienced similar effects as the top-floor apartment except for the roof related parameters (Table Appendix C.4). Furthermore, the sensitivity index of the HSPT parameter is further reduced for the mid-floor apartment due to the additional assumption of an adiabatic roof. This assumption reduces the convective heat transfer to the surroundings due to reduced surface area in contact with the external environment.

Appendix D. BIC Feature Selection Results

BIC Results for the Detached Archetype: The BIC analysis of the detached dwelling involves top five models for energy consumption parameters (Table Appendix D.1). Based on BIC scores, Model 1 is the 'best model'. However, the BIC weights do not significantly favour one model over another and $\triangle$BIC suggests that each model holds a substantial level of empirical support.

When comparing these results to sensitivity rankings, Model 5 includes the floor U-value and exclude the roof emissivity parameter in the list of influential parameters. However, Roof_E is consistently ranked above UFloor in both Morris and Sobol sensitivity rankings. Therefore, model 5 can be removed from consideration. The same is true for Model 4, which includes the external door U-value instead of UFloor and UDoor is also consistently ranked below Roof_E. Therefore, Model 4 can also be removed from consideration. Models 1 and 2 conform with Morris rankings but not with Sobol rankings (Table Appendix D.2). Also, One of the aims behind this study is add some objectivity to decisions regarding input data assumptions and simplifications that ultimately lead to an increase in modeling accuracy and/or decrease in modeling time. Therefore, model 3, which comprises a minimum set of parameters, is chosen as the final energy consumption model.

Similarly, when observing BIC results for the detached heating demand (Table Appendix D.1), the model weights or △BIC do not favour either model subset from the top three models. The Morris sensitivity results suggest that all three models are viable. However, the Sobol sensitivity results indicate that only Model C is correct since Roof_E is ranked below UFrame for this model. It was found in the literature that the variance-based methods such as the Sobol method are known to provide more reliable results than other sensitivity measures [16]. Therefore, it is assumed that the rankings produced by the Sobol method are more accurate/reliable than those by the Morris method. Henceforth, model C is chosen as the final BIC heating load model for this archetype.

BIC Results for the Semi-detached Archetype: Based on the energy consumption. all top model subsets include the lower ranked wall thermal emissivity parameter and excludes the higher ranked Roof_E parameter (Table Appendix D.2). Nevertheless, the BIC evaluation metrics favour Model 1, which carries the greatest weight and lowest BIC score. The values of △BIC indicate that Models 2 and 3 still possess considerable empirical evidence to be considered as the optimum model subset. However, the selected parameters in Models 2 and 3 disagree with Morris and Sobol rankings. Model 2 includes the lower ranked UFloor parameter and excludes the higher ranked Roof_E parameter. Model 3 excludes the higher ranked building orientation parameter and includes the lower ranked the auxiliary energy parameter. Hence, model 1 is chosen as the final BIC energy consumption model.

The heating demand BIC results are more definitive than those obtained for the energy consumption (Table Appendix D.2). Model A has a significantly higher weight compared to Models B and C. Also, the △BIC for Model C is close to the range whereby this model can be completely discarded. Model B still possess considerable empirical evidence as a possible model subset as per the △BIC value. However, given the significantly higher BIC weight, Model A is chosen as the final BIC heating demand model.

BIC Results for the Top-floor Apartment: The energy consumption BIC results for this archetype indicate that Models 1, 2 and 3 comprise the most probable optimum subset (Table Appendix D.3). However, Model 2 does not conform with Sobol sensitivity rankings. The model subset includes the window frame U-value parameter only and excludes higher ranked parameters such as Wall_abs, Wall_E and HSBT. Both Models 1 and 3 agree with the sensitivity rankings except for the building orientation parameter which is ranked as one of the top ten parameters in both sensitivity methods. While BIC metrics appear to favour the top three models, they all exhibit some conflicting characteristics compared to the sensitivity analysis results. Hence, lower weighted models are assessed for their suitability. Model 4 is identical to Model 1 except this model includes the building orientation parameter, which is deemed significant in the sensitivity results. Despite the BIC weight of Model 4 being considerably smaller than the weights given to the top three models, this model is chosen as the final BIC energy consumption model due to the conformity with the sensitivity results. Additionally, the analysis excludes Model 5 as this model has the lowest BIC weight and conflicts with the sensitivity results.

The heating demand BIC results strongly suggest that both Model A and Model B comprise the optimum model subset as per the BIC weights (Table Appendix D.4). While Model A attains a higher weight, the $\triangle$BIC for Model B indicates that this could be the 'true' model. Also, Model A conflicts with the sensitivity results as this model includes the UFrame parameter and excludes the more influential HSBT parameter. Therefore, Model B is chosen as the final BIC heating demand model given the model's conformity with sensitivity rankings and favorable BIC evaluation metrics. Additionally, the heating load results validate the selection of Model 4 for energy consumption parameters since all BIC heating load models incluse the orientation parameters in the influential parameter list. A possible reason for the exclusion of the building orientation parameter in the energy consumption results could be due to high variance caused by the heating system COP parameter. This high fluctuation in the output along with the non-linear and non-monotonic nature of the orientation parameter could have excluded this parameter from the optimum model subsets.

BIC Results for the Mid-floor Apartment: The mid-floor energy consumption BIC results clearly illustrate that Model 1 is the optimum model subset with a substantial BIC weight compared to Models 2 and 3 (Table Appendix D.5). When analyzing the $\triangle$BIC values, Model 2 could be considered a 'true' model. However, this model includes the window visible transmittance parameter in th parameter subset suggest. The transmittance parameter exhibits zero/negligible effect on the output variance as per SA results. Although Model 3 conforms with the sensitivity rankings, this model could be excluded on the account of the significantly low weight and higher degree of model complexity compared to Model. Hence, Model 1 is chosen as the final BIC energy consumption model.

The mid-floor heating demand BIC results provide more definitive results compared to other archetype models (Table Appendix D.5). Model A receives a significantly higher weight of 86% when compared to other models. The parameter subset of Model A also conforms well with sensitivity results. Model A is chosen as the final BIC heating load model as this model offers reduced complexity and has favorable BIC metrics.

Table Appendix B.1: Morris Rankings for Energy Consumption, Heating & Electrical Load of Detached and Semi-detached Archetypes.

Detached Archetype				Semi Detached			
Rank	Energy	Heating Load	Electrical Load	Rank	Energy	Heating Load	Electrical Load
1	COP	HSPT	Equip	1	COP	HSPT	Equip
2	HSPT	Occ	Ld	2	HSPT	ACR	Ld
3	Occ	ACR	Aux	3	Occ	Occ	Aux
4	UWall	UWall	-	4	ACR	Equip	-
5	ACR	Equip	-	5	UGlaze	UGlaze	-
6	Equip	UGlaze	-	6	Equip	UWall	-
7	UGlaze	SHGC	-	7	UWall	SHGC	-
8	SHGC	WWR	-	8	SHGC	WWR	-
9	Ld	URoof	-	9	Ld	Ld	-
10	URoof	Ld	-	10	WWR	URoof	-
11	WWR	Wall_abs	-	11	URoof	DHW	-
12	DHW	DHW	-	12	DHW	Wall_abs	-
13	Wall_abs	Roof_abs	-	13	Orientation	Orientation	-
14	Aux	Wall_abs	-	14	Wall_abs	Roof_abs	-
15	Roof_abs	Roof_E	-	15	Aux	Roof_E	-
16	Roof_E	UDoor	-	16	Roof_abs	Wall_E	-
17	Wall_E	UFloor	-	17	Roof_E	UFloor	-
18	UDoor	UFrame	-	18	Wall_E	UDoor	-
19	UFloor	Orientation	-	19	UFloor	HSBT	-
20	UFrame	HSBT	-	20	UDoor	UFrame	-
21	Orientation	UPart	-	21	HSBT	UPart	-
22	UPart	Vt	-	22	UFrame	Vt	-
23	HSBT	-	-	23	UPart	-	-
24	Vt	-	-	24	Vt	-	-

Table Appendix B.2: Morris Rankings for Energy Consumption, Heating & Electrical Load of Top and Mid-floor Apartments.

Top-floor Apartment				Mid-floor Apartment			
Rank	Energy	Heating Load	Electrical Load	Rank	Energy	Electrical Load	Electrical Load
1	COP	ACR	Equip	1	COP	Equip	Equip
2	ACR	Occ	Ld	2	ACR	Ld	Ld
3	Occ	HSPT	Aux	3	Occ	Aux	Aux
4	HSPT	URoof	-	4	HSPT	-	-
5	URoof	Equip	-	5	UGlaze	-	-
6	Orientation	Orientation	-	6	Equip	-	-
7	Equip	UGlaze	-	7	UWall	-	-
8	UGlaze	SHGC	-	8	SHGC	-	-
9	SHGC	DHW	-	9	Orientation	-	-
10	DHW	UWall	-	10	DHW	-	-
11	UWall	Ld	-	11	WWR	-	-
12	WWR	WWR	-	12	Ld	-	-
13	Ld	Roof_abs	-	13	Aux	-	-
14	Roof_abs	Roof_E	-	14	Wall_abs	-	-
15	Roof_E	Wall_abs	-	15	Wall_E	-	-
16	Aux	Wall_E	-	16	UFrame	-	-
17	Wall_abs	UFrame	-	17	HSBT	-	-
18	Wall_E	HSBT	-	18	UPart	-	-
19	UFrame	UPart	-	19	Vt	-	-
20	HSBT	Vt	-				
21	UPart	-	-				
22	Vt	-	-				

Table Appendix C.1: Detached Sobol Sensitivity Indices (%) & Rankings.

Energy Consumption (%)			Heating Load (%)			Electrical Load (%)	
1st Order Effect	Total Order Effect	Rank	1st Order Effect	Total Order Effect	Rank	Total Order Effect	Rank
53.8917	58.8890	COP	29.9423	30.7196	HSPT	90.64	Equip
12.5264	14.3101	HSPT	19.9765	20.3345	Occ	9.02	Ld
8.3504	9.7794	Occ	15.3060	15.5926	ACR	0.35	Aux
6.4682	7.4258	ACR	10.2093	11.0083	Equip	0	COP
3.8032	4.4347	UWall	10.1137	10.3613	UWall	0	HSPT
3.3326	4.0467	Equip	6.4421	7.1897	UGlaze	0	ACR
2.7143	2.9301	UGlaze	3.6836	4.1371	SHGC	0	Occ
1.4418	1.9071	SHGC	0.6894	1.7999	WWR	0	UGlaze
0.3991	0.9470	WWR	0.8970	0.9133	Ld	0	UWall
0.3794	0.4930	URoof	0.8871	0.9059	URoof	0	SHGC
0.3543	0.4548	Ld	0.2627	0.3214	Wall_abs	0	WWR
0.1036	0.2044	Wall_abs	0.2344	0.2346	DHW	0	URoof
0.0261	0.1125	DHW	0.0978	0.1615	Wall_E	0	DHW
0.0366	0.0489	Wall_E	0.0727	0.1027	Roof_abs	0	Wall_abs
0.0484	0.0484	Aux	0.0685	0.0705	UFrame	0	Roof_E
0.0347	0.0368	Roof_abs	0.0323	0.0623	UFloor	0	Roof_abs
0.0219	0.0330	UFrame	0.0487	0.0615	Roof_E	0	Wall_E
0.0145	0.0323	Roof_E	0.0166	0.0291	Orientation	0	UDoor
0.0059	0.0297	UFloor	0.0191	0.0271	UDoor	0	UFloor
0.0073	0.0131	UDoor	0.0020	0.0026	HSBT	0	UFrame
0.0032	0.0100	Orientation	0.0007	0.0008	UPart	0	Orientation
0.0067	0.0084	HSBT	0	0	Vt	0	UPart
0.0014	0.0017	UPart	0	0	Aux	0	HSBT
0	0	Vt	0	0	COP	0	Vt

Table Appendix C.2: Semi-detached Sobol Sensitivity Indices (%) & Rankings.

Energy Consumption (%)			Heating Load (%)			Electrical Load (%)	
1st Order Effect	Total Order Effect	Rank	1st Order Effect	Total Order Effect	Rank	Total Order Effect	Rank
58.1842	62.6856	COP	30.2067	30.9379	HSPT	88.81	Equip
11.6271	13.2716	HSPT	20.7285	21.0245	ACR	10.82	Ld
8.0127	9.1533	ACR	14.3440	14.5233	Occ	0.35	Aux
5.5314	6.4704	Occ	10.3860	11.0241	Equip	0	COP
3.3343	3.6297	UGlaze	8.6610	9.5953	UGlaze	0	HSPT
2.8592	3.4494	Equip	5.9672	6.1631	UWall	0	ACR
1.9400	2.4292	SHGC	5.3420	5.8963	SHGC	0	Occ
2.0713	2.4268	UWall	0.9002	2.2451	WWR	0	UGlaze
0.4541	1.0311	WWR	1.1620	1.1736	Ld	0	UWall
0.3582	0.4722	Ld	0.6445	0.6716	URoof	0	SHGC
0.2495	0.3296	URoof	0.3547	0.3547	DHW	0	WWR
0.0528	0.1183	Wall_abs	0.0883	0.2043	Orientation	0	URoof
0.0288	0.1013	Roof_abs	0.1695	0.2027	Wall_abs	0	DHW
0.0510	0.0610	DHW	0.0582	0.1046	Wall_E	0	Wall_abs
0.0179	0.0587	Orientation	0.0576	0.087	Roof_abs	0	Roof_E
0.0429	0.0429	Aux	0.0360	0.064	UFloor	0	Roof_abs
0.0087	0.0272	Roof_E	0.0401	0.0555	Roof_E	0	Wall_E
0.0077	0.0268	UFloor	0.0143	0.0146	UFrame	0	UDoor
0.0174	0.0263	Wall_E	0.0076	0.0118	UDoor	0	UFloor
0.0032	0.0062	UFrame	0.0018	0.0032	UPart	0	UFrame
0.0017	0.0061	UDoor	0.0023	0.0026	HSBT	0	Orientation
0.0058	0.0058	HSBT	0	0	Vt	0	UPart
0.0007	0.0025	UPart	0	0	Aux	0	HSBT
0	0	Vt	0	0	COP	0	Vt

Table Appendix C.3: Top-floor Apartment Sobol Sensitivity Indices (%) & Rankings.

Energy Consumption (%)			Heating Load (%)			Electrical Load (%)	
1st Order Effect	Total Order Effect	Rank	1st Order Effect	Total Order Effect	Rank	Total Order Effect	Rank
57.2784	62.8408	COP	25.2409	25.6607	ACR	86.76	Equip
9.8760	11.3301	ACR	21.2714	22.1296	HSPT	13.17	Ld
8.2282	9.7499	HSPT	19.5417	19.8734	Occ	0.36	Aux
7.2764	8.6824	Occ	8.2731	8.6935	URoof	0	COP
3.1314	3.3440	URoof	7.0719	7.3233	Equip	0	HSPT
2.5157	2.8588	Equip	3.8269	4.2383	UGlaze	0	ACR
1.4607	1.8705	UGlaze	2.8086	3.4747	SHGC	0	Occ
1.1618	1.5771	Orientation	2.9433	3.4058	Orientation	0	UGlaze
1.1398	1.4700	SHGC	2.1972	2.1972	DHW	0	UWall
0.7099	1.0078	DHW	2.0700	2.1897	UWall	0	SHGC
0.7377	0.9481	UWall	1.0744	1.1064	Ld	0	WWR
0.3628	0.4971	Ld	0.3122	0.7833	WWR	0	URoof
0.0184	0.3865	WWR	0.3639	0.4266	Roof_abs	0	DHW
0.0982	0.1596	Roof_E	0.2650	0.3197	Roof_E	0	Wall_abs
0.1076	0.1368	Roof_abs	0.0792	0.1119	Wall_abs	0	Roof_E
0.0410	0.0410	Aux	0.0087	0.0289	Wall_E	0	Roof_abs
0.0221	0.0373	Wall_abs	0.0052	0.0196	HSBT	0	Wall_E
0.0206	0.0280	Wall_E	0.0024	0.0091	UFrame	0	UDoor
0.0061	0.0117	HSBT	0.0041	0.0051	UPart	0	UFloor
0.0011	0.0097	UFrame	0	0	Vt	0	UFrame
0.0020	0.0043	UPart	0	0	COP	0	Orientation
0	0	Vt	0	0	Aux	0	UPart

Table Appendix C.4: Mid-floor Apartment Sobol Sensitivity Indices (%) & Rankings.

Energy Consumption (%)			Heating Load (%)			Electrical Load (%)	
1st Order Effect	Total Order Effect	Rank	1st Order Effect	Total Order Effect	Rank	Total Order Effect	Rank
47.3168	53.2272	COP	31.8914	32.8976	ACR	87.48	Equip
15.0071	17.2667	ACR	20.7707	21.4567	Occ	12.95	Ld
9.5664	10.8438	Occ	16.8237	18.1642	HSPT	0.36	Aux
8.4510	9.5054	HSPT	8.1231	8.5859	Equip	0	COP
3.7839	4.5807	Equip	5.3921	6.0352	UGlaze	0	HSPT
2.5813	3.2661	UGlaze	3.4242	4.1817	SHGC	0	ACR
1.6315	2.2727	SHGC	3.4946	3.9671	Orientation	0	Occ
1.7822	2.1522	Orientation	3.0754	3.2928	UWall	0	UGlaze
1.5263	1.7108	UWall	2.6322	2.6329	DHW	0	UWall
1.2381	1.4443	DHW	0.9642	1.8264	WWR	0	SHGC
0.4124	1.0230	WWR	1.1097	1.1934	Ld	0	WWR
0.6512	0.7759	Ld	0.0937	0.0937	Wall_abs	0	URoof
0.0795	0.0796	Aux	0.0218	0.0288	Wall_E	0	DHW
0.0373	0.0476	Wall_abs	0.0088	0.0110	UFrame	0	Wall_abs
0.0159	0.0293	Wall_E	0.0079	0.0094	UPart	0	Roof_E
0.0035	0.0118	HSBT	0.0024	0.0046	HSBT	0	Roof_abs
0.0066	0.0093	UFrame	0	0	Vt	0	Wall_E
0.0025	0.0062	UPart	0	0	COP	0	UDoor
0	0	Vt	0	0	Aux	0	UFloor

Table Appendix D.1: Detached Archetype BIC Results for the Energy Consumption and Heating Demand

Energy Consumption						Heating Demand			
Parameter	Model 1	Model 2	Model 3	Model 4	Model 5	Parameter	Model A	Model B	Model C
COP	✓	✓	✓	✓	✓	HSPT	✓	✓	✓
HSPT	✓	✓	✓	✓	✓	Occ	✓	✓	✓
Occ	✓	✓	✓	✓	✓	ACR	✓	✓	✓
ACR	✓	✓	✓	✓	✓	Equip	✓	✓	✓
UWall	✓	✓	✓	✓	✓	UWall	✓	✓	✓
Equip	✓	✓	✓	✓	✓	UGlaze	✓	✓	✓
UGlaze	✓	✓	✓	✓	✓	SHGC	✓	✓	✓
SHGC	✓	✓	✓	✓	✓	WWR	✓	✓	✓
WWR	✓	✓	✓	✓	✓	Ld	✓	✓	✓
URoof	✓	✓	✓	✓	✓	URoof	✓	✓	✓
Ld	✓	✓	✓	✓	✓	Wall_abs	✓	✓	✓
Wall_abs	✓	✓	✓	✓	✓	DHW	✓	✓	✓
DHW	✓	✓	✓	✓	✓	Wall_E	✓	✓	✓
Wall_E	✓	✓	✓	✓	✓	Roof_abs	✓	✓	✓
Aux	✓	✓	✓	✓	✓	UFrame	✗	✗	✗
Roof_abs	✓	✓	✓	✓	✓	UFloor	✗	✗	✗
UFrame	✗	✗	✗	✗	✗	Roof_E	✓	✓	✗
Roof_E	✓	✓	✗	✗	✗	Orientation	✗	✗	✗
UFloor	✗	✗	✗	✗	✓	UDoor	✓	✗	✗
UDoor	✓	✗	✗	✓	✗	HSBT	✗	✗	✗
Orientation	✗	✗	✗	✗	✗	UPart	✗	✗	✗
HSBT	✗	✗	✗	✗	✗	Vt	✗	✗	✗
UPart	✗	✗	✗	✗	✗				
Vt	✗	✗	✗	✗	✗				
BIC Score	160,094.8	160,094.9	160,095.3	160,095.4	160,096.6	BIC Score	160,094.8	160,094.9	160,095.3
Weight	0.239	0.233	0.190	0.182	0.097	Weight	0.239	0.233	0.190
$\triangle BIC$	0	0.052	0.460	0.551	1.181	$\triangle BIC$	0	0.052	0.460
No. of Parameters	18	17	16	17	17	No. of Parameters	18	17	16

Table Appendix D.2: Semi-detached Archetype BIC Results for the Energy Consumption and Heating Demand

Energy Consumption				Heating Demand			
Parameter	Model 1	Model 2	Model 3	Parameter	Model A	Model B	Model C
COP	✓	✓	✓	HSPT	✓	✓	✓
HSPT	✓	✓	✓	ACR	✓	✓	✓
ACR	✓	✓	✓	Occ	✓	✓	✓
Occ	✓	✓	✓	Equip	✓	✓	✓
UGlaze	✓	✓	✓	UGlaze	✓	✓	✓
Equip	✓	✓	✓	UWall	✓	✓	✓
SHGC	✓	✓	✓	SHGC	✓	✓	✓
UWall	✓	✓	✓	WWR	✓	✓	✓
WWR	✓	✓	✓	Ld	✓	✓	✓
Ld	✓	✓	✓	URoof	✓	✓	✓
URoof	✓	✓	✓	DHW	✓	✓	✓
Wall_abs	✓	✓	✓	Orientation	✓	✓	✓
Roof_abs	✓	✓	✓	Wall_abs	✓	✓	✓
DHW	✓	✓	✓	Wall_E	✓	✓	✓
Orientation	✓	✓	✗	Roof_abs	✓	✓	✓
Aux	✓	✓	✓	UFloor	✓	✓	✓
Roof_E	✗	✗	✗	Roof_E	✓	✓	✗
UFloor	✗	✓	✗	UDoor	✓	✓	✓
Wall_E	✓	✓	✓	UFrame	✗	✗	✗
UFrame	✗	✗	✗	UPart	✗	✗	✗
UDoor	✗	✗	✗	HSBT	✗	✓	✗
HSBT	✗	✗	✗	Vt	✗	✗	✓
UPart	✗	✗	✗				
Vt	✗	✗	✗				
BIC Score	156,654.8	156,658.7	156,659.2	BIC Score	136,051.5	136,055.8	136,060.7
Weight	0.69	0.10	0.08	Weight	0.87	0.10	0.009
△BIC	0	3.89	4.34	△BIC	0	4.35	9.18
No. of Parameters	17	18	16	No. of Parameters	18	19	19

Table Appendix D.3: Top-floor Apartment BIC Results for the Energy Consumption.

Parameter	Model 1	Model 2	Model 3	Model 4	Model 5
COP	✓	✓	✓	✓	✓
ACR	✓	✓	✓	✓	✓
HSPT	✓	✓	✓	✓	✓
Occ	✓	✓	✓	✓	✓
URoof	✓	✓	✓	✓	✓
Equip	✓	✓	✓	✓	✓
UGlaze	✓	✓	✓	✓	✓
Orientation	✗	✗	✗	✓	✗
SHGC	✓	✓	✓	✓	✓
DHW	✓	✓	✓	✓	✓
UWall	✓	✓	✓	✓	✓
Ld	✓	✓	✓	✓	✓
WWR	✓	✓	✓	✓	✓
Roof_E	✓	✓	✓	✓	✓
Roof_abs	✓	✓	✓	✓	✓
Aux	✓	✓	✓	✓	✓
Wall_abs	✗	✗	✓	✗	✗
Wall_E	✗	✗	✗	✗	✗
HSBT	✗	✗	✗	✗	✓
UFrame	✗	✓	✗	✗	✗
UPart	✗	✗	✗	✗	✗
Vt	✗	✗	✗	✗	✗
BIC Score	143,906.9	143,908.1	143,909.1	143,911.1	143,912.0
Weight	0.45	0.24	0.15	0.07	0.04
△BIC	0	1.21	2.18	4.15	5.11
No. of Parameters	15	16	16	16	16

Table Appendix D.4: Top-floor Apartment BIC Results for the Heating Demand

Parameter	Model A	Model B	Model C	Model D	Model E
ACR	✓	✓	✓	✓	✓
HSPT	✓	✓	✓	✓	✓
Occ	✓	✓	✓	✓	✓
URoof	✓	✓	✓	✓	✓
Equip	✓	✓	✓	✓	✓
UGlaze	✓	✓	✓	✓	✓
SHGC	✓	✓	✓	✓	✓
Orientation	✗	✗	✗	✓	✗
DHW	✓	✓	✓	✓	✓
UWall	✓	✓	✓	✓	✓
Ld	✓	✓	✓	✓	✓
WWR	✓	✓	✓	✓	✓
Roof_abs	✓	✓	✓	✓	✓
Roof_E	✓	✓	✓	✓	✓
Wall_abs	✓	✓	✓	✓	✓
Wall_E	✓	✓	✓	✓	✓
HSBT	✗	✗	✗	✗	✗
UFrame	✓	✗	✓	✓	✗
UPart	✗	✗	✓	✗	✓
Vt	✗	✗	✗	✗	✗
BIC Score	128,125.8	128,126.3	128,132.6	128,133.0	128,133.3
Weight	0.53	0.40	0.017	0.014	0.012
$\triangle$BIC	0	0.55	6.84	7.27	7.50
No. of Parameters	17	16	18	18	17

Table Appendix D.5: Mid-floor Apartment BIC Results for the Energy Consumption and Heating Demand

Energy Consumption				Heating Demand			
Parameter	Model 1	Model 2	Model 3	Parameter	Model A	Model B	Model C
COP	✓	✓	✓	ACR	✓	✓	✓
ACR	✓	✓	✓	Occ	✓	✓	✓
Occ	✓	✓	✓	HSPT	✓	✓	✓
HSPT	✓	✓	✓	Equip	✓	✓	✓
Equip	✓	✓	✓	UGlaze	✓	✓	✓
UGlaze	✓	✓	✓	SHGC	✓	✓	✓
SHGC	✓	✓	✓	Orientation	✓	✓	✓
Orientation	✓	✓	✓	UWall	✓	✓	✓
UWall	✓	✓	✓	DHW	✓	✓	✓
DHW	✓	✓	✓	WWR	✓	✓	✓
WWR	✓	✓	✓	Ld	✓	✓	✓
Ld	✓	✓	✓	Wall_abs	✓	✓	✓
Aux	✓	✓	✓	Wall_E	✓	✓	✓
Wall_abs	✓	✓	✓	UFrame	✗	✓	✗
Wall_E	✗	✗	✓	UPart	✗	✗	✗
HSBT	✗	✗	✗	HSBT	✗	✗	✗
UFrame	✗	✗	✗	Vt	✗	✗	✓
UPart	✗	✗	✗				
Vt	✗	✓	✗				
BIC Score	140,278.1	140,281.9	140,282.1	BIC Score	128,439.7	128,444.4	128,446.8
Weight	0.64	0.093	0.087	Weight	0.86	0.08	0.02
△BIC	0	3.85	4.00	△BIC	0	4.6	7.1
No. of Parameters	14	15	15	No. of Parameters	13	14	14