buildings

MDPI

Article

Development of Building Design Optimization Methodology: Residential Building Applications

Yeonjin Bae [1,*,†], Donghun Kim [2] and William Travis Horton [3]

1 Building Technologies Research and Integration Center, Oak Ridge National Laboratory, Oak Ridge, TN 37830, USA
2 Building Technology and Urban System Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA
3 School of Civil Engineering, Purdue University, West Lafayette, IN 47907, USA
* Correspondence: baey@ornl.gov
† Yeonjin's work was done while she was at Purdue University.

Abstract: Building design optimization is a highly complex problem, requiring long computational running processes because of the many options that exist when a building is being designed. This paper introduces an integrated approach through which to perform this optimization within an acceptable time frame. The approach includes the methods of variable selection, model simplification, and a sequential optimization process. Using singular value decomposition, a large number of design variables is reduced to a smaller subset that can be solved more quickly through the optimization algorithm. To expedite the variable selection process, a modeling approach that quickly simulates annual energy consumption was developed to replace full annual energy simulations. The developed methodology was applied to two residential buildings in the US, and the results are discussed herein. To assess the accuracy of the integrated optimization methodology, the optimized life cycle costs are compaa variables demonstrating the strongest contributions in the optimization study were identified. The proposed methodology significantly shortened the time requirements for the optimization processes of the two case studies by 74% and 84%; the optimized life cycle costs were within 0.05% and 0.06%, respectively, of the optimum point.

Keywords: building design optimization; energy simulation; variable selection; life cycle cost

check for updates

Citation: Bae, Y.; Kim, D.; Horton, W.T. Development of Building Design Optimization Methodology: Residential Building Applications. *Buildings* **2024**, *14*, 107. https://doi.org/10.3390/buildings14010107

Academic Editor: Danny Hin Wa Li

Received: 25 October 2023
Revised: 14 November 2023
Accepted: 20 November 2023
Published: 31 December 2023

1. Introduction

According to energy usage data from the US Energy Information Administration, residential and commercial buildings contribute approximately 40% of energy consumption and 16% of energy-related carbon emissions in the US [1,2]. Over the past 25 years, residential energy use has increased by approximately 7.0%, and that of commercial buildings has increased by about 10.4% [3]. As energy consumption in the US continues to increase and as energy-related carbon emissions need to be reduced, energy simulation tools have been incorporated more regularly into the building design process for high-performance buildings. Energy-efficient design optimization techniques for buildings are undoubtedly important to reducing building energy consumption and its associated costs, which have been active research topics [4].

Ideally, to find an optimal building design point that minimizes an objective function, such as energy consumption or building life cycle cost (LCC), building energy simulation software, such as EnergyPlus [5] and TRNSYS [6], can be coupled with an optimization algorithm for accurate energy consumption calculations. However, such optimization studies are usually prohibitive because of their very long calculation times, owing to large numbers of design variables, and their relatively long simulation times related to the software used to predict annual energy consumption at each iteration of the optimization process.

A lot of research has been carried out on reducing the long computational times associated with the optimization process. The primary focus of previous research has been to develop simplified energy consumption models that can replace detailed energy simulation software using, for example, neural networks and regression models. Magnier and Haghighat [7], as well as Gossard et al. [8], used a genetic algorithm with an artificial neural network model to optimize thermal comfort and energy consumption in a residential house. Ghiaus [9] applied robust regression (an alternative of the least squares regression method that is effective when outliers exist) to predict the heating load as a function of outdoor temperature. Catalina et al. [10] used a quadratic polynomial regression model to predict heating demand with multiple independent variables, such as shape factor, building time constant, and more. To generate the database for creating the model, 18,144 simulations were performed. Hygh et al. [11] used EnergyPlus to build a multivariate regression model with 27 parameters; 20,000 full annual simulation datasets were generated to predict heating and cooling loads. Assuming each simulation takes 30 s to 1 min, the process of generating the dataset alone requires 166 to 332 h. The cooling regression model has a high value for the coefficient of determination (R^2) that indicates how well the data fit the model; however, the heating regression model has a low value for R^2 from 0.498 to 0.816. Ghiaus [9] used the concept of balance point temperature to construct a regression model. Hygh et al. [11] used a fixed 18 °C balance temperature to predict energy consumption, and Krarti [12] estimated the exact balance temperature. Eisenhower [13] developed an analytical metamodel that fit the building simulation data and then performed the optimization. Although a simplified modeling approach is capable of carrying out the building design optimization process with little computational effort, it essentially requires a tremendous number of pre-energy simulation results to obtain a reliable model, which is the major disadvantage of this approach. Consequently, regenerating the database is necessary when changing the design variables to create a simplified model.

Another approach through which to accelerate the optimization process involves employing variable selection methods to choose a subset of variables from the entire set of variables. Bettonvil [14] applied a group screening method to detect important factors by applying sequential bifurcation to a building energy model. Rahni [15] partitioned building parameters into multiple groups and tested which groups make more significant contributions to building energy performance. Corrado and Mechri [16] used the Morris method to reduce the parameter dimensions from 129 to 10. Brohus et al. [17] used an analysis of variance-based analysis to identify significant variables of residential building energy consumption, and identified the 10 most important parameters out of a total of 57.

An alternative approach is using sensitivity analysis. In previous studies, sensitivity analyses were employed in either the pre- or postoptimization phases or to efficiently select a subset of variables significantly affecting the objective function, thus streamlining the optimal design process with reduced effort compared with full optimization. Gunay et al. [18] used sensitivity analysis to identify the key operational parameters of office buildings, and Cheng et al. [19] used the Morris one-at-a-time method to design a sustainable housing community. Østergård et al. [20] used sensitivity analysis to inform the decision-making process in the early building design process. Sensitivity analysis is easy to use and is a valuable tool for assessing model behavior and identifying critical input parameters; therefore, it is frequently used in building simulation research. However, sensitivity analysis becomes more challenging as the dimensionality of the input space increases, and the interactions between variables can become complex, making it difficult to isolate the effect of an individual variable [21].

To speed up building design optimization, the two approaches of simplified modeling and variable reduction can naturally be combined. However, despite many previous studies, it is hard to find studies testing the integrated optimization approach. Additionally, the computational time needed to generate a reduced-order model or a variable selection process is often neglected. Therefore, it is not clear how much the integrated approach is able to accelerate building design optimization.

This paper presents an integrated optimization methodology to tackle the building design optimization problem. The developed methodology can be used for any residential building design optimization process, including for the early design and engineering design stages. Also, the developed methodology is not limited to any specific design variables for the optimization process. A novel variable selection method is proposed and adopted for the optimization problem to reduce the high dimensionality of the design variables. To overcome the long computational time required to generate sufficient data that can be used during the variable selection process, a new, simplified modeling approach is developed to replace the full annual energy simulation for residential buildings. A strategy through which to approach the true optimal point is also introduced. The integrated methodology is applied to two case studies of typical residential buildings in the US, and comparisons of the results with those of the full optimization processes for the entire design spaces are presented.

2. Overall Methodology

To derive an accurate but computationally efficient methodology for building design optimizations, two main problems need to be solved: (1) reduce the number of optimization variables and (2) reduce the amount of computational time consumed by the energy simulation software. To overcome these problems, the following process shown in Figure 1 is proposed.

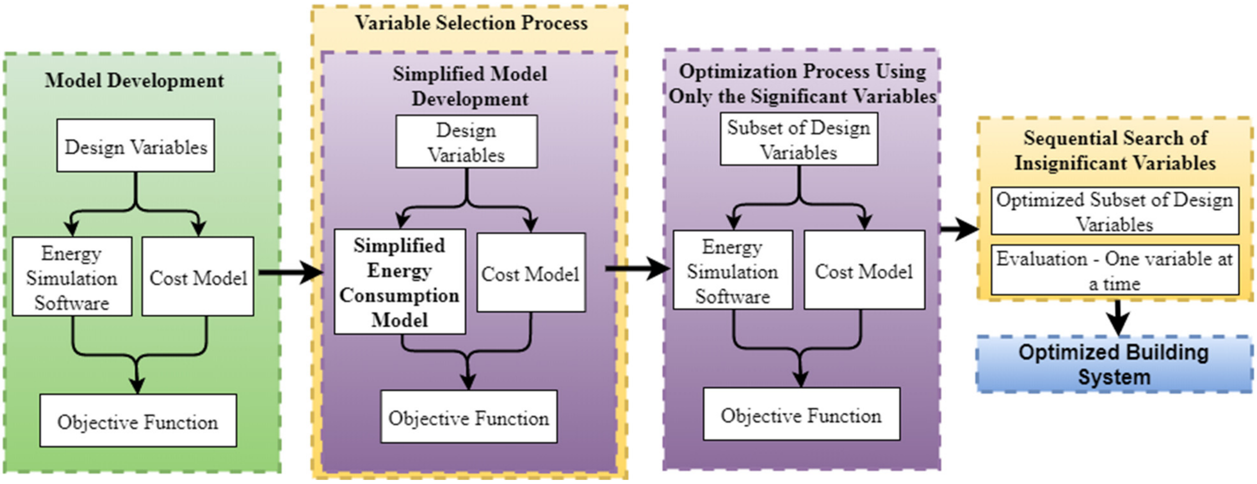


Figure 1. Building optimization process.

The first step, shown in the green box, is to define a building design optimization problem by specifying appropriate design variables and their corresponding constraints, along with the objective function (e.g., the minimization of LCC for residential buildings in this paper). To evaluate energy costs accurately during the optimization process, detailed building energy simulation software is incorporated rather than the simplified models that are typically used in other research approaches. The second step is to reduce the number of design variables using a variable selection method, which identifies subvariables that have a significant influence on the objective function. To accelerate this process, which requires a large number of annual building energy simulations, a simplified modeling strategy is adopted. Then, an optimization algorithm is coupled to the building energy software, which searches for the optimum point only in a reduced dimension of the design space as defined by the variable selection method. In this approach, the simplified energy consumption models are only used during the variable selection process (the second block of Figure 1), and high-fidelity models are used in the actual optimization process (the third block of Figure 1). The last step is to reduce the optimization error that is inherited from the approximation approach (i.e., the variable selection algorithm) by sequentially adjusting the insignificant variables.

3. Variable Selection Algorithm and Optimization Process

In this section, each step shown in Section 2 is described in detail. Section 3.1 provides the details of the variable selection algorithm, and Section 3.2 shows how this variable selection algorithm is applied to an optimization problem. Section 3.3 shows the simplified energy consumption model used for residential building applications that will be used only in the variable selection process, and Section 3.4 shows the optimization process, with the variables selected from the variable selection process and postoptimization process.

3.1. Variable Selection Algorithm Using Singular Value Decomposition

Consider a differentiable function, denoted as $f : (x_1, x_2, \cdots, x_n) \mapsto y$, where $y \in \mathcal{R}$, $x_i \in \mathcal{D}_i, \forall i \in \{1, 2, \cdots, n\}$, and $\mathcal{D}_i \subset \mathcal{R}$ is a feasible set for a variable x_i. The task here is to select the important subvariables out of $(x_1, x_2, \cdots, x_n)$, which influence y most significantly. Selecting a metric that quantifies the magnitude of importance or significance of a variable is the key step. We define the following operator L:

$$L(v) := \frac{1}{\iint_{\mathcal{D}} dx} \iint_{\mathcal{D}} | < \nabla f(x), v > |^2 dx, \tag{1}$$

where $\mathcal{D} = \mathcal{D}_1 \times \cdots \times \mathcal{D}_n$, $\nabla f(x_1, x_2, \cdots, x_n) = \sum_{i=1}^{n} e_i \frac{\partial f}{\partial x_i}(x_1, x_2, \cdots, x_n)$, e_i is the natural basis corresponding to x_i, $< \cdot, \cdot >$ is the inner product, and $v (\in \mathcal{R}^n)$ is a unit vector. Operator L measures the spatial average of the directional derivate of f over $\mathcal{D}$ in the direction of $\boldsymbol{v}$. When $L(\boldsymbol{v}) = 0$, y does not change in the direction of v. When $L(v_1) >> L(v_2)$, y varies more significantly in the direction of v_1 than that of v_2 in the spatial average sense. Therefore, employing L as a measure to gauge the significance or importance of a variable is a rational choice.

The following example is provided for further clarity. Let $y = {x_1}^2 + 0.01{x_2}^2$, and assume that x_1 and x_2 are well-scaled such that $-1 \leq x_1, x_2 \leq 1$. Variable x_1 clearly influences the output variable y more significantly than it does x_2. The L values for $e_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $e_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$, which correspond to x_1 and x_2, respectively, are

$$L(e_1) = \iint_{-1 \leq x_i \leq 1} < \begin{bmatrix} 2x_1 \\ 0.02x_2 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \end{bmatrix} >^2 dx_1 dx_2 = \iint_{-1 \leq x_i \leq 1} 2^2 {x_1}^2 dx_1 dx_2 = 2^2 \times \frac{2}{3} \times 2 \text{ and}$$

$$L(e_2) = \iint_{-1 \leq x_i \leq 1} < \begin{bmatrix} 2x_1 \\ 0.02x_2 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} >^2 dx_1 dx_2 = \iint_{-1 \leq x_i \leq 1} 0.02^2 {x_2}^2 dx_1 dx_2 = 0.02^2 \times \frac{2}{3} \times 2.$$

For these L values, $L(e_1) >> L(e_2)$. This example shows that the operator L provides a measure of the contribution of a variable to an output variable.

For practical applications, Equation (1) can be approximated as

$$L(v) \approx \frac{1}{N} \sum_{k=1}^{N} \Big| < \nabla f(x_k), v > |^2, \tag{2}$$

where N is a total number of samples of x_k over a feasible domain $\mathcal{D} = \mathcal{D}_1 \times \cdots \times \mathcal{D}_n$. With this metric, the most important variable is represented by

$$\underset{||v||=1}{\arg\max}\, L(v). \tag{3}$$

The maximizer can be obtained by performing singular value decomposition (SVD) of the matrix JJ^T, where $J = \begin{bmatrix} \nabla f(\boldsymbol{x}_1) & \nabla f(\boldsymbol{x}_2) & \cdots & \nabla f(\boldsymbol{x}_N) \end{bmatrix} \left(\in \mathcal{R}^{n \times N}\right)$. We denote the resulting form of the singular value decomposition as

$$JJ^T = V\Sigma V^T = \begin{bmatrix} V_1 & V_2 \end{bmatrix} \begin{bmatrix} \Sigma_1 & \\ & \Sigma_2 \end{bmatrix} \begin{bmatrix} V_1^T \\ V_2^T \end{bmatrix}, \tag{4}$$

where $V_1(\in \mathcal{R}^{n\times r}), V_2(\in \mathcal{R}^{n\times(n-r)}), \Sigma_1 \in \mathcal{R}^{r\times r}, \Sigma_2 \in \mathcal{R}^{(n-r)\times(n-r)}$ are submatrices of V and Σ, and

$$\Sigma_1 = \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \end{bmatrix}, \Sigma_2 = \begin{bmatrix} \sigma_{r+1} & & \\ & \ddots & \\ & & \sigma_n \end{bmatrix}.$$

The diagonal matrix Σ contains magnitudes of L in descending order corresponding to each column vector of V. The first column vector of V maximizes the metric L (i.e., the maximizer of Equation (3)), and the second column vector of V maximizes the metric L over a space perpendicular to the first column vector of V. Consequently, the last column vector has a minimum value of L. Therefore, primary coordinates can be selected out of V in which y varies significantly by looking at the values of Σ; these coordinates can denote V_1 as the primary component. The criteria can be set to select V_1 so that the cumulative sum of the percentage of Σ_1 is at least 99.5%, which means that 99.5% of the magnitudes of L can be explained by primary component V_1, and Σ_2 is negligible [15].

3.2. Application to an Optimization Problem

Consider an optimization problem,

$$\begin{aligned} &\min_{x\in\mathcal{R}^n} f(\boldsymbol{x}) \\ &s.t. A\boldsymbol{x} \leq b, \end{aligned} \tag{5}$$

where f is a differentiable function for $\forall x \in \mathcal{R}^n$. We define new variables of z such that $z = V^T x$. The ith element of z is the projection of x onto the ith column vector of V. Note that x can be expressed as $x = V_1 z_1 + V_2 z_2$ because $V = \begin{bmatrix} V_1 & V_2 \end{bmatrix}$ and V is orthonormal. Because $L(V_1) >> L(V_2)$, the optimization problem can be searched over a subspace spanned by the column vectors of V_1 rather than over the entire $\mathcal{R}^n$ space. In other words, neglecting the influence of z_2 in the optimization problem becomes possible because the magnitudes of the directional derivatives along the column vector of V_2 are small. The new optimization problem in a lower dimension then becomes

$$\begin{aligned} &\min_{z_1\in\mathcal{R}^r} f(V_1 z_1) \\ &s.t. (AV_1) z_1 \leq b. \end{aligned} \tag{6}$$

For some cases, it is worthwhile to retrieve the important variables in the original design space from z_1 because z_1 is a linear combination of design variables; thus, z_1 may not have a direct physical meaning. The significant design variables in the original space can be retrieved by examining the column vectors of V_1. For example, when the jth component of a column vector of V_1 is more predominant than the others, it means that the jth component of x contributes to the output significantly. The retrieving process may increase the dimension of the design space from r to $m(\geq r)$.

The final optimization problem with the selected m variables in the original design space becomes

$$\begin{aligned} &\min_{x\in E} f(\boldsymbol{x}) \\ &s.t. A\boldsymbol{x} \leq b, \end{aligned} \tag{7}$$

where $E \subset R^m$.

The proposed method is very similar to that of principle component analysis (PCA). PCA is also a dimensionality reduction technique commonly used in statistic and machine learning. The method aims to reduce the data's dimensionality while retaining as much relevant information as possible [22]. However, it differs significantly because (1) the matrix to be composed by SVD contains information of the output variable, and PCA does not, and (2) the proposed method is developed from deterministic points of view, and PCA is based on random variables. It also differs from the partial least squares method (PLS). PLS is a multivariate statistical technique used for modeling the relationship between a set of independent variables and a set of dependent variables [23]. PLS is commonly used in regression analysis and classification tasks to establish predictive models, but it assumes a linear relationship between input and output variables. In contrast, SVD does not inherently assume linearity. Additionally and more importantly, our proposed method can be applied to an optimization process as described here, and PCA and PLS cannot [24,25].

3.3. Simplified Energy Consumption Modeling Approach

The introduced variable selection algorithm requires calculating the Jacobian matrix of the objective function at multiple points in a design space (see Equation (4)). This method would take a tremendous time with approximately (n variables) $\times$ (N design points) $\times$ (T annual simulation times per a design point). To reduce the time required for generating the series of Jacobian matrices, the modified degree day method is proposed. The degree day method is easy to implement and requires only basic weather data, but it has limitations because it assumes a linear relationship between temperature and energy use, and other factors such as insulation level, building design, and occupant behavior can also significantly affect energy use. Given the robust correlation between degree day data and energy consumption in residential buildings, and coupled with the relatively limited influence of occupant behavior, the degree day method continues to be a valuable and straightforward tool for energy modeling in the residential sector. The proposed method is based on the following assumptions and modifications to the existing degree day method, and the detailed process is shown in Figure 2:

- The actual degree day varies with the chosen design variables. Consequently, the balance temperature must be calculated for each design combination;
- The modified ANAGRAM method is used to shift the diagonal linear regression models to the origin based on daily average degree days to separately predict heating, cooling, and fan energy consumption;
- To find an appropriate number of data samples to yield an accurate regression model, the recursive least square (RLS) algorithm is used;
- A simplified model is constructed specifically to replace energy simulation software during the variable selection phase. This model is employed for the sole purpose of identifying which variables significantly influence the output. Therefore, the model must be sufficiently accurate to discern these influential trends.

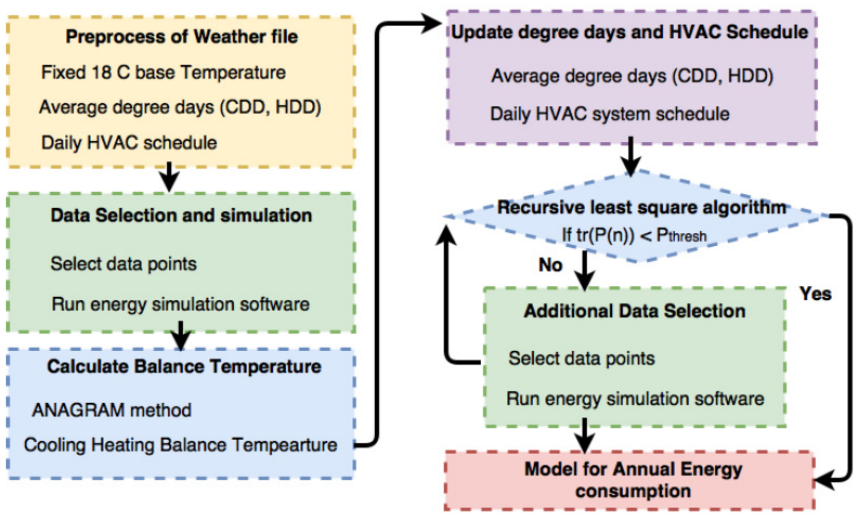


Figure 2. Process of the simplified energy consumption model's development.

Suppose we want to evaluate annual energy consumption y for a residential building at a design point x, which consists of building design parameters. The role of building an energy simulation model, denoted as f, that maps x to y, is $y = f(x, w)$, where w represents any other inputs that influence y, such as weather data. The goal is to find a function that approximates f. We adopt the degree day method for its simplicity and widespread use in predicting energy consumption for a long time period for residential buildings. More precisely, it can be represented as the following relationship over a heating season, for example:

$$\begin{aligned} y_P &\approx \alpha \times DD_P \\ DD_P &= \int_{t_0}^{t_0+P} (T_\mathrm{b} - T_\mathrm{a}(\tau))d\tau, \end{aligned} \tag{8}$$

where y_P is the heating energy consumption for the period of $[t_0, t_0 + P]$, and α is a building-specific coefficient. The variable DD_P is the number of heating degree days for the period, and $(T_\mathrm{a}, T_\mathrm{b})$ are the ambient and balance temperatures of the building, respectively. Note that once α and T_b are identified, the annual energy consumption for heating can be easily predicted by only considering the weather data. In other words, when α and T_b are known, annual heating energy consumption can be predicted without simulating a detailed building energy model. The time-invariant parameters of α and T_b depend on $\boldsymbol{x}$. To estimate them, a regression approach is used with samples of y_P obtained from building energy simulations for selected days. The final form of our simplified model is

$$y_P \approx \hat{\alpha}(x) \int_{t_0}^{t_0+P} (\hat{T}_\mathrm{b} - T_\mathrm{a}(\tau))d\tau, \tag{9}$$

where $\hat{\alpha}$ and $\hat{T}_\mathrm{b}$ are the estimated values of α and T_b at a design point of $\boldsymbol{x}$. After the simulation studies for residential buildings, a day is selected as P (i.e., $P = 1$ day). A natural question regards the selection of an appropriate sample size to accurately estimate the parameters; that is, the question of how many daily simulations are needed to accurately estimate α and T_b. This problem can be handled by employing the RLS method because it calculates confidence intervals for estimated parameters as each sample is updated. Because the original model, Equation (8), is a nonlinear function with respect to the parameters, it was reformulated as follows:

$$y_P = \theta_1 DD_{P,18} + \theta_2, \tag{10}$$

where $\theta_1 = \alpha$, and $\theta_2 = \alpha P(T_\mathrm{b} - 18)$. The equation $DD_{P,18} = \int_{t_0}^{t_0+P} (18 - T_\mathrm{a}(\tau))d\tau$ represents the number of degree days with the assumption of an 18 °C balance temperature [9].

The model structure, shown in Equation (10), can be readily used with the RLS method because of its linearity. Once the parameters of θ_1 and θ_2 are estimated, the balance temperature can be retrieved by $T_\mathrm{b} = 18 + \frac{1}{P}\frac{\theta_2}{\theta_1}$.

As an initial dataset, 6 days from the weather file are selected and simulated; these include the heating design day as well as 5 other days, which are uniformly distributed and are found by preprocessing a typical meteorological year file at the location. Next, θ_1 and θ_2 are calculated from the six samples. Then, they are inserted into the RLS method as the initial guess. Another day simulation is implemented, and those parameters are updated. This process is repeated until the parameter covariance goes below a threshold. For a description of the RLS method, refer to Appendix A. Because heating energy, cooling energy, and fan energy consumption need to be considered to predict total annual energy consumption, separate models for heating, cooling, and fan energies are developed using the same methodology as used here. For brevity, in this paper, the equations for cooling and fan energy are omitted.

To calculate $\partial f / \partial x_i(x)$, two simplified models at two design points of $(x, x + \Delta x_i)$ are constructed. The process is repeated for each variable and each data point to construct a series of Jacobian matrices (shown in Equation (4)).

3.4. Optimization with the Significant Variables and Sequential Search Method

After identifying significant variables using the variable selection method with the simplified modeling approach, an optimization routine can be implemented with those variables while fixing the remaining variables, which are insignificant, to reference values. In the current case study, a discrete binary version of the particle swarm optimization (PSO) methodology [26] is adopted. This algorithm is selected based on its ability to efficiently explore the design space and arrive at an optimal solution to overcome the problems of the continuous version that cannot handle discrete variables. To prevent fast convergence to a local optimum, the inertia PSO version [27] is used. The results that are discussed in Section 4 demonstrate that it is possible to get close to the optimum point when using only those significant variables that were identified using the variable selection method. Once an optimal point is found with the significant variables, there may still be a chance to get closer to the true optimum point by perturbing the insignificant variables, although their contributions are expected to be small. One point of clarification is that the variables that are identified as insignificant are not necessarily unimportant; they just do not demonstrate a significant contribution within the scope of the optimization study, which is why they are excluded in that phase. However, to further improve the final result, a sequential one-dimensional optimization approach is used. This approach optimizes the first variable in the group of insignificant variables while fixing the others. Once the process is terminated, the second variable in that group is optimized while fixing those that remain. This process is repeated until the last variable is optimized. The order of the sequential search approach may be important because variables are sorted according to their contributions to the cost function. Therefore, the order of variables in the sequential search approach follows that determined in the variable selection process.

4. Case Studies

To validate the developed methodology, two cases studies for residential buildings were selected. Two representative residential building types in the US were chosen, and design variables were selected from typical variables that are often considered in the engineering design stage.

4.1. Descriptions of Case Study Buildings

To develop detailed building models, the energy simulation software EnergyPlus was used. The buildings are a typical slab-on-grade residential house and a house with a heated basement. They were referenced from the Pacific Northwest National Laboratory residential prototype [28] and were assumed to be located in Indianapolis, Indiana. Figure 3 shows the model dimensions of the houses.

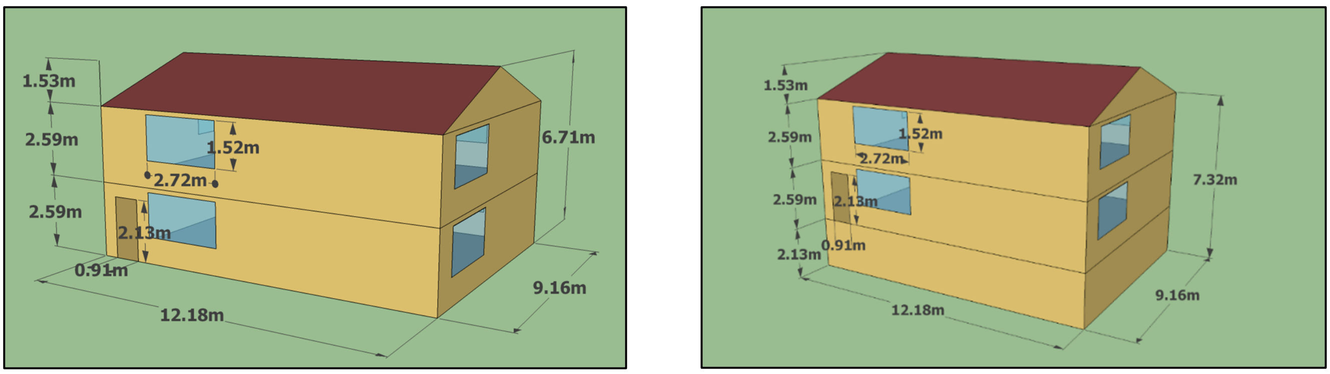


Figure 3. Dimensions of the case study houses.

The first house has a gross floor area of 334.6 m^2 with two different spaces, which include a living space and attic. The living space is the only conditioned zone, and the net conditioned area is 223.1 m^2. The entry doors are located on the south and north sides of the building, and two windows are placed on each side of the building. The house is assumed to have three bedrooms and three bathrooms. The second house has a gross floor area of 446.1 m^2. The house has a living space, attic, and heated basement. The prototype complies with the 2012 International Energy Conservation Code with modified construction layers for optimization purposes. The living space is the only conditioned zone, and the net conditioned area is 334.6 m^2. The house has four bedrooms and four bathrooms, and the basement is placed completely below the ground level.

4.2. Defining Design Variables for Optimization

To compare the results of the proposed method with those of the full optimization, we chose a relatively small number of design variables in this paper. The following considerations were made in defining the design variables for the case studies: select elements that may have (1) a high impact on building energy consumption, (2) a strong influence in the construction cost, and (3) energy-saving or cost-saving benefits, but whose influence is not strictly known beforehand. Based on these considerations, 12 different construction design variables were selected to investigate their effect on the LCC of the residential building. The first eight variables were related to the building envelope, and the remaining variables were associated with the heating, ventilation, and air-conditioning (HVAC) system.

Commercially available products associated with those design variables are listed in Table 1. For example, for roofing material, one can take either asphalt shingles, a metal surface, or a concrete tile roof. Design variables of the wall core can be a given stud dimension and filled with any type of insulation shown in the table. Structural insulated panels or insulated concrete forms with various thicknesses were also considered to be different wall core materials for this study. The under-floor insulation was used only for case study 1, and the foundation wall insulation was used only for case study 2. The total number of possible combinations, $\prod_{i=1}^{n} N_i$, is around 60 million for each case, where N_i is the number of options for the ith variable. Notably, it can be reduced significantly to $\prod_{i=1}^{r} N_i$ through variable selection, because the reduced dimension r is generally much smaller than the original dimension n. For the case studies, $n = 12$, $r = 5$ (see the later discussion), and, thus, the total number of combinations was reduced to around 700.

Table 1. Design variables and values.

Design Variables (Number of Options)	Available Values
Roofing Material (3)	Asphalt shingles, metal surface, concrete tile roof
Roof Eave Overhang Depth (3)	305 mm (12 in.), 457 mm (18 in.), 610 mm (24 in.)
Attic Insulation Material (12)	Loose fill cellulose: R3.3 (IP [1]-R19), R4.4 (IP-R25), R5.3 (IP-R30), R6.7 (IP-R38), R8.6 (IP-R49), R10.6 (IP-R60)
	Fiberglass batting: R3.3 (IP-R19), R4.4 (IP-R25), R5.3 (IP-R30), R6.7 (IP-R38), R8.6 (IP-R49), R10.6 (IP-R60)
External Wall Siding Material (4)	Vinyl siding, wood siding, fiber cement siding, brick
Wall Core (16)	Stud: 38 × 89 mm (2 × 4 in.) studs at 400 mm (16 in.), 38 × 140 mm (2 × 6 in.) studs at 600 mm (24 in.) on center, 38 × 184 mm (2 × 8 in.) studs at 600 mm (24 in.) on center Insulation: filled with fiberglass batting insulation, sprayed-on foam insulation, loose-fill cellulose insulation
	Structural insulated panels: 114 mm (3 5/8 in.), 165 mm (5 5/8 in.), 210 mm (7 3/8 in.), 260 mm (9 3/8 in.)
	Insulated concrete forms: 228 mm (9 in.), 278 mm (11 in.), 328 mm (13 in.)

Table 1. *Cont.*

Design Variables (Number of Options)	Available Values
External Foam Board Insulation (6)	Board insulation: 12.7 mm (0.5 in.), 25.4 mm (1 in.), 38.1 mm (1.5 in.), 50.8 mm (2 in.), 63.5 mm (2.5 in.), 76.2 mm (3 in.)
Under-Floor Insulation or Foundation Wall Insulation (8)	Extruded polystyrene: 25.4 mm (1 in.), 50.8 mm (2 in.), 76.2 mm (3 in.), 101.6 mm (4 in.)
	Expanded polystyrene: 25.4 mm (1 in.), 50.8 mm (2 in.), 76.2 mm (3 in.), 101.6 mm (4 in.)
Window Type (2)	Double-pane window, triple-pane window
Air Conditioner Speed (2)	Single-speed, multispeed
Heat Recovery Type (2)	None, sensible heat recovery
Seasonal Coefficient of Performance (Air Conditioner SEER [2]) (6)	3.81 (SEER 13), 4.10 (SEER 14), 4.40 (SEER 15), 4.69 (SEER 16), 4.98 (SEER 17), 5.28 (SEER18)
Natural Gas Furnace Efficiency (4)	80%, 85%, 90%, 95%

[1] IP stands for imperial units; [2] SEER represents Seasonal Energy Efficiency Ratio.

4.3. Optimization Objective Function

We considered an LCC over a 20-year time horizon as the objective function, as shown in Equation (11). To estimate realistic construction costs, the material cost, labor, overhead, and profit were all considered. The primary tool used for estimating construction costs is RSMeans [29]. RSMeans is a construction cost database that is widely used and respected in the United States. It provides comprehensive, up-to-date, and reliable information related to construction costs, materials, labor, and equipment cost. To account for regional variations in construction costs, a location influence factor (I_f) is used, and the I_f of 92.5% (applicable for Indianapolis) is multiplied by the national average cost. HVAC equipment costs were modeled using multiple linear regression to fit cost data taken from online equipment suppliers relative to both system capacity and efficiency. The discount rate was assumed to be 3%, and the modified uniform present value (UPV^*) factors were taken from the annual supplement to the NIST Handbook [30]. The UPV^* factors are energy price projections for different fuel types based on the US Department of Energy/US Energy Information Administration regional projections.

$$LCC = I_f C_{\text{Con}} + C_{\text{HVAC}} + C_{\text{Elec}} UPV^*_{\text{Elec}} + C_{\text{NG}} UPV^*_{\text{NG}} \quad (11)$$

Here, the following pertains:

LCC—incremental LCC of given building system;
I_f—influence factor of the location;
C_{Con}—construction cost;
C_{HVAC}—HVAC equipment cost;
C_{Elec}—electricity cost;
C_{NG}—natural gas cost;
UPV^*_{Elec}—UPV* factor for electricity cost;
UPV^*_{NG}—UPV* factor for natural gas cost.

5. Case Study Results

5.1. Validation Methodology

To validate and compare the accuracy and efficiency of the proposed methodology, a full optimization with all design variables was first performed. The full optimization was intended to minimize the LCC using all 12 variables. Following the full optimization study, the proposed methodology was implemented. The required data size, N in Equation (4), may be automatically selected by tracking the convergence behavior of singular values as

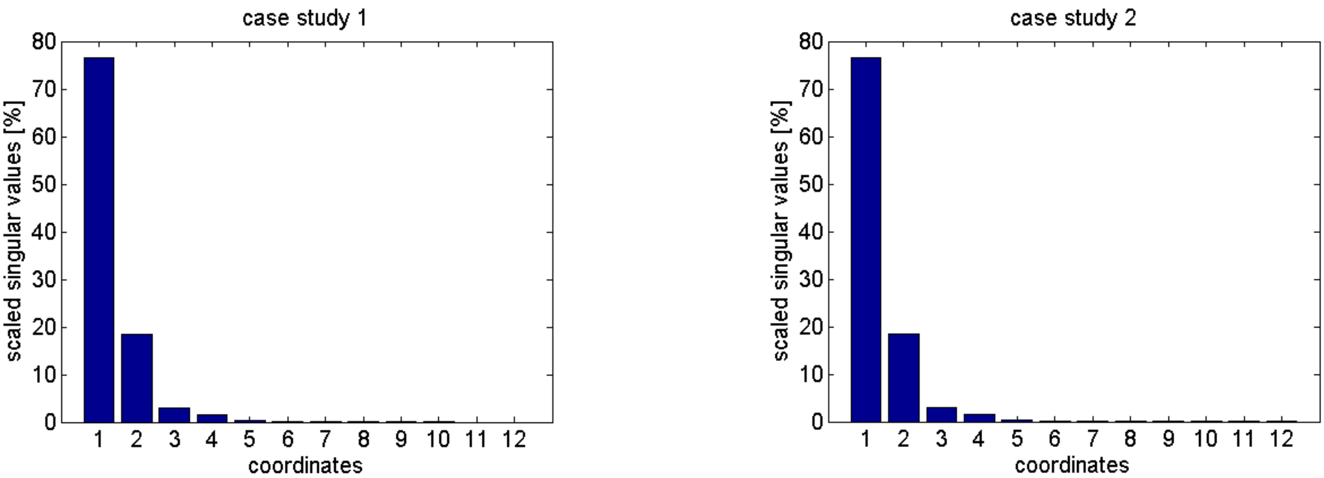
N increases. However, for the feasibility study of the proposed method, a fixed size of data (i.e., 100 data points) was generated to identify significant variables.

After finding the significant variables using the variable selection method, the remaining variables were fixed to the cheapest material, and the optimization process was carried out on the significant variables. For more detailed information about the initial setting of variables, refer to Appendix B. Finally, the best values of the remaining insignificant variables were determined using a sequential search method.

The inertia weight version of the binary PSO algorithm was used for both full and proposed optimizations. Based on the suggestions of Parsopoulos and Vrahatis [31], the swarm size for the PSO algorithm and the neighborhood size of a von Neumann topology were set as 64 and 8, respectively, for the full optimization, and were 36 and 6, respectively, for the proposed method.

5.2. Variable Selection Results

Figure 4 shows the singular values of the Gramian matrix (see Equation (4)) in descending order. The singular values are scaled with respect to the sum of all singular values. For both case studies, the first four coordinates are selected based on the fact that their cumulative sum of singular values is at least 99.49% and 99.46% with respect to the sum of all singular values. In other words, using only the first four coordinates, 99.49% and 99.46% of data can be captured for case studies 1 and 2. Because the variables are discrete, it is convenient to find the significant parameters in the original design space using Equation (7).



(**a**) (**b**)

Figure 4. (**a**) Result of singular values in case study 1. (**b**) Result of singular values in case study 2.

Table 2 displays the singular vectors of the first four coordinates with the higher values shown in bold, indicating the variables that significantly contribute to the corresponding coordinates. For example, external wall siding material and wall core type make the strongest contributions to the first and second singular vectors, and roofing materials, air-conditioning (AC) speed, and heat recovery type are the primary contributors to the third singular vector. Through this process, 5 out of the 12 design variables, including roofing material, external wall siding material, wall core, AC speed, and heat recovery, were identified as significant variables in both case studies. Although the same five design variables are selected for both case studies, this choice does not necessarily imply that these five design variables are significant variables for all cases. The results will vary based on the specified design parameter values and regional factors, including construction cost and equipment cost, as well as weather conditions.

Table 2. Singular vectors. [3]

Case Study Building 1	Coordinate 1	Coordinate 2	Coordinate 3	Coordinate 4
Roofing Material	−0.012	0.024	0.401	−0.910
Roof Eave Overhang Depth	0.000	−0.001	−0.030	0.003
Attic Insulation Material	−0.006	−0.006	0.014	0.101
External Wall Siding Material	**0.999**	0.028	0.017	−0.005
External Foam Board	0.000	−0.005	−0.041	−0.010
Wall Core	−0.028	**0.999**	−0.008	0.023
Foundation Wall Insulation	−0.003	−0.009	−0.021	−0.069
Window Type	−0.002	0.001	0.125	0.053
AC Speed	−0.004	−0.001	**0.274**	0.118
Heat Recovery Type	−0.012	−0.002	**0.847**	**0.367**
Air Conditioner Seasonal Coefficient of Performance	−0.001	0.000	0.041	0.018
Natural Gas Furnace Efficiency	−0.003	−0.002	0.163	0.071
Case Study Building 2	**Coordinate 1**	**Coordinate 2**	**Coordinate 3**	**Coordinate 4**
Roofing material	−0.012	0.024	**0.416**	**−0.903**
Roof Eave Overhang Depth	0.000	−0.001	−0.030	0.003
Attic Insulation Material	−0.006	−0.006	0.013	0.101
External Wall Siding Material	**0.999**	0.028	0.017	−0.004
External Foam Board	0.000	−0.006	−0.041	−0.011
Wall Core	−0.028	**0.999**	−0.009	0.023
Under-floor Wall Insulation	−0.002	−0.009	−0.023	−0.074
Window Type	−0.002	0.001	0.130	0.057
AC Speed	−0.004	−0.001	**0.271**	0.122
Heat Recovery Type	−0.012	−0.002	**0.837**	**0.379**
Air Conditioner Seasonal Coefficient of Performance	−0.001	0.000	0.040	0.018
Natural Gas Furnace Efficiency	−0.003	−0.002	0.176	0.080

[3] Significant variables in both case studies are highlighted in blue in the first column.

5.3. Simplified Energy Consumption Model

To ensure that the simplified modeling approach is sufficiently accurate for the purpose of the variable selection, the variable selection algorithm is also implemented using the full annual energy model for comparison. Then, the first four singular vectors obtained from the simplified model, denoted as $V_s = [v_{s,1}, v_{s,2}, v_{s,3}, v_{s,4}]$, and from the full model, denoted as $V_f = [v_{f,1}, v_{f,2}, v_{f,3}, v_{f,4}]$, are compared. Table 3 shows the matrix of $V_s^T V_f$ for the case 1 and 2 buildings. If the simplified modeling approach is accurate, ${v_{s,i}}^T v_{f,j} \approx \pm\delta_{i,j}$ where $\delta_{i,j}$ is the Kronecker delta. That is, the diagonal elements of $V_s^T V_f$ must be close to either 1 or −1, and the off-diagonal elements must be close to 0. Notably, the 0 property arises because of the orthonormal characteristic of singular vectors. The results satisfy the criteria, which demonstrates that the resulting significant variables from the simplified modeling approach are very similar to those of the full model for both case studies.

Table 3. Correlation matrices between singular vectors obtained from detailed and simplified models [4].

Case 1				Case 2			
−1.00	−0.08	−0.02	0.02	−0.99	−0.11	−0.05	0.02
−0.08	0.99	0.05	0.04	−0.11	0.99	−0.01	−0.04
−0.02	−0.05	1.00	−0.07	0.05	0.00	−0.99	0.14
−0.02	0.04	−0.07	−0.99	−0.01	−0.04	−0.14	−0.97

[4] Diagonal elements of $V_s^T V_f$ are shown in grey.

5.4. Optimization Result Comparison

To compare the accuracy of the integrated optimization methodology, the optimized LCCs were compared with the full optimizations, and they are shown in Table 4. Because it is of interest to see the improvement resulting from incorporating the sequential search method, the optimization results with only the selected variables (i.e., without the sequential method) were also included.

Table 4. Comparison of optimized LCC. [5]

Case Study Building 1	Optimized LCC (USD)	Percentage Difference (%)
Optimization with the Significant Variables	53,062	4.19
Optimization with the Significant Variables and Sequential Search	50,956	0.05
Optimization with All Design Variables	50,929	-
Case Study Building 2	**Optimized LCC (USD)**	**Percentage Difference (%)**
Optimization with the Significant Variables	57,818	4.60
Optimization with the Significant Variables and S Search	55,308	0.06
Optimization with All Design Variables	55,273	-

[5] This cost does not include construction and HVAC costs in Equation (11) that are not associated with design variables.

The results show that the minimal LCC errors arising from only using the subset of design variables were around 4% for both case studies. By further employing the sequential search method, the errors were brought to within 0.05% (USD 27) and 0.06% (USD 35) for a 20-year horizon.

Table 5 shows a comparison of the optimized designs using full optimization and the proposed methodology. In case study 1, the difference was caused only by the roof eave overhang depth and attic insulation material. For case study 2, roof eave overhang depth and air conditioner seasonal energy efficiency ratio (SEER) rating caused the difference. As expected from the variable selection algorithm, the variables are insignificant in both cases.

All numerical experiments were run on a desktop computer with a 6 core 3.10 GHz CPU, 6 GB RAM, and Windows (64-bit) operating systems. For both cases, the variable selection process took a majority of the total optimization time. However, the variable selection process without introducing the simplified modeling approach took 13.5 for case study building 1 and 9.98 h for case study building 2. This time demonstrates the efficiency of the proposed modeling methodology. The full optimization process required approximately 33 and 29 h, and the proposed methodology required 8.7 and 5.7 h for each case. In other words, the proposed optimization approach achieved 74% and 84% computational time savings. These significant reductions were mainly because of the reduced design space and the number of evaluations requested by the optimization algorithm.

Table 5. Comparison of optimal design. [6]

Case Study Building 1	Optimal Design with Full Optimization	Optimal Design with the Proposed Method
Roofing Material	F12 asphalt shingles	F12 asphalt shingles
Roof Eave Overhang Depth	305 mm (12 in.)	457 mm (18 in.)
Attic Insulation Material	Attic loose fill—R3.3 (IP-R19)	Attic loose fill—R4.4 (IP-R25)
External Wall Siding Material	F11 wood siding	F11 wood siding
External Foam Board	12.7 mm (0.5 in.)	12.7 mm (0.5 in.)
Wall Core	38 × 140 mm (2 × 6 in.) studs at 600 mm (24 in.) on center filled with loose fill cellulose insulation	38 × 140 mm (2 × 6 in.) studs at 600 mm (24 in.) on center filled with loose fill cellulose insulation

Table 5. *Cont.*

Case Study Building 1	Optimal Design with Full Optimization	Optimal Design with the Proposed Method
Foundation Wall Insulation	Expanded polystyrene 50.8 mm (2 in.)	Expanded polystyrene 50.8 mm (2 in.)
Window Type	Triple-pane window	Triple-pane window
AC Speed	Multispeed	Multispeed
Heat Recovery Type	Sensible	Sensible
Seasonal Coefficient of Performance (Air Conditioner SEER)	5.28 (SEER 18)	5.28 (SEER 18)
Natural Gas Furnace Efficiency	0.95	0.95
Case Study Building 2	**Optimal Design with Full Optimization**	**Optimal Design with the Proposed Method**
Roofing Material	F12 asphalt shingles	F12 asphalt shingles
Roof Eave Overhang Depth	305 mm (12 in.)	457 mm (18 in.)
Attic Insulation Material	Attic loose fill—R4.4 (IP-R25)	Attic loose fill—R4.4 (IP-R25)
External Wall Siding Material	F11 wood siding	F11 wood siding
External Foam Board	12.7 mm (0.5 in.)	12.7 mm (0.5 in.)
Wall Core	38 × 140 mm (2 × 6 in.) studs at 600 mm (24 in.) on center filled with loose fill cellulose insulation	38 × 140 mm (2 × 6 in.) studs at 600 mm (24 in.) on center filled with loose fill cellulose insulation
Under-Floor Insulation	Expanded polystyrene 25.4 mm (1 in.)	Expanded polystyrene 25.4 mm (1 in.)
Window Type	Triple-pane window	Triple-pane window
AC Speed	Multispeed	Multispeed
Heat Recovery Type	Sensible	Sensible
Seasonal Coefficient of Performance (Air Conditioner SEER)	4.98 (SEER 17)	5.28 (SEER 18)
Natural Gas Furnace Efficiency	0.95	0.95

[6] The differences are shown in blue rows in the table.

6. Conclusions

To make building design optimization feasible, a novel integrated approach is presented in this paper. This work includes a method to reduce the dimensionality of a design space, which aims to identify only a few significant variables that need to be considered as optimization variables. The method formally defines a measure of significance and uses the SVD to order variables based on the metric using sample data. A simplified modeling methodology to reduce the time needed to generate samples for the variable selection process is introduced. The developed methodology is applied to case studies for residential buildings, and the results are compared with those of the full optimization process over the entire design space.

In summary:

- Using the variable selection process, significant variables (5 out of 12) that demonstrate the strongest contribution to the optimization study are identified;
- The proposed methodology significantly shortens the time requirement for the optimization process in the two case studies of 74% and 84%, and the optimized LCC is within 0.05% and 0.06%, respectively, of the optimum point.

These findings not only validate the effectiveness of the integrated approach, but also highlight its potential to streamline and expedite the design optimization process while maintaining a high degree of accuracy in the results.

Author Contributions: Y.B.: Model development and simulation, data analysis, writing—original draft, writing—review and editing. D.K.: model development, writing—original draft. W.T.H.:

project administration, writing—review and editing, supervision. All authors have read and agreed to the published version of the manuscript.

Funding: This work was partially funded by field work proposal CEBT105 under DOE BTO activity no. BT0302000 and BT0305000.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to privacy.

Acknowledgments: This manuscript was authored by UT-Battelle LLC under contract DEAC05-00OR22725 with DOE. The publisher, by accepting the article for publication, acknowledges that the US government retains a nonexclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this manuscript, or allow others to do so, for US government purposes. DOE will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan (https://www.energy.gov/doe-public-access-plan (accessed on 18 September 2023)).

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A

To find an appropriate number of data samples to yield an accurate regression model, the regression least square (RLS) algorithm is used. The RLS gives an advantage related to computational efficiency by immediately acquiring coefficients of a_j with the current data sample without storing past observation data [32]. In addition to this advantage, the required number of data samples can be determined based on the confidence of parameters.

Suppose a true model has the following general linear form:

$$y(k) = a_0 + a_1 x_1(k) + a_2 x_2(k) + \cdots a_m x_m(k) + e(k), \tag{A1}$$

where $y(k) \in \mathcal{R}$ and $x(k) \in \mathcal{R}^m$ are the outputs and inputs, respectively. Variable $e(k) \in \mathcal{R}$ is a mean zero random variable. The coefficients of a_j are the true parameters. Equation (A1) can be rewritten as follows:

$$y(k) = \phi^T(k)\theta_T + e(k). \tag{A2}$$

Let θ be candidate parameters for θ_T, and define model error as $\epsilon(k;\theta) := y(k) - \phi^T(k)\theta$. For a single-output system, the least squares estimator tries to minimize $\frac{1}{n}\sum_{k=1}^{n} \epsilon\left(k;\theta\right)^2$. In other words, the least square problem is $\min_{\theta} \frac{1}{n}\sum_{k=1}^{n}\left(y(k) - \phi^T(k)\theta\right)^2$, where n is the number of data points. Let the optimal solution with data length n be $\hat{\theta}_n$. The least squares solution is given by

$$\hat{\theta}_n = (\frac{1}{n}\sum_{k=1}^{n}\phi(k-1)\phi(k-1)^T)^{-1}\frac{1}{n}\sum_{k=1}^{n}\phi(k-1)y(k).$$

The RLS algorithm tries to find $\hat{\theta}_{n+1}$ from $\hat{\theta}_n$ iteratively without re-solving the optimization problem as the data size increases. The iteration is given by [32,33].

$$\hat{\theta}(n) = \hat{\theta}(n-1) + K(n-1)(y(t) - \phi(n-1)^T\hat{\theta}(n-1)), \tag{A3}$$

$$P(n-1) = P(n-2) - K(n-1)\phi(n-1)^T P(n-2), \text{ and} \tag{A4}$$

$$K(n-1) = P(n-2)\phi(n-1) \times (1 + \phi(n-1)^T P(n-2)\phi(n-1))^{-1}. \tag{A5}$$

Here, the notation was switched to $\hat{\theta}(n)$ rather than $\hat{\theta}_n$, which is the least square solution. This switch is intended to distinguish the solutions of the least square and the RLS because of an initial condition of the RLS problem. A remark can made on P in Equation (A4). It is well-known that $P(n-1)$ is proportional to $E(\tilde{\theta}_n\tilde{\theta}_n^T)$, where $\tilde{\theta}_n = \theta_T - \hat{\theta}_n$, which is the covariance of the error associated with estimated parameters. In other words, it is the uncertainty of parameters estimated with data size of n. Therefore, by tracking $tr(P(n-1))$, the required data size can be determined. More precisely, the optimal data size, denoted as n_{opt}, is determined by the following criteria:

$$n_{\text{opt}} = \min\{n | tr(P(n)) \le p_{\text{thresh}}, n \in \mathcal{N}\}, \quad \text{(A6)}$$

where p_{thresh} is the threshold for determining the number of the data size of n. In this paper, we set the initial condition of P, P_0 as 10^5 and p_{thresh} as 10^{-1} based on a simulation study.

Appendix B

Table A1 displays the initial variable values for both the full optimization and optimization with the significant variables for case study buildings 1 and 2. Variables that are not included in the optimization with significant values are set to the cheapest material, and these variables are determined via the sequential search approach.

Table A1. Initial variable values used for optimization.

Case Study Building 1	**Full Optimization**	**Optimization with the Significant Variables**
Roofing Material	F12 asphalt shingles	F12 asphalt shingles
Roof Eave Overhang Depth	305 mm (12 in.)	-
Attic Insulation Material	Attic loose fill—R3.3 (IP-R19)	-
External Wall Siding Material	Vinyl siding	Vinyl siding
External Foam Board	12.7 mm (0.5 in.)	-
Wall Core	38 × 89 mm (2 × 4 in.) studs at 400 mm (16 in.) on center filled with loose fill cellulose insulation	38 × 89 mm (2 × 4 in.) studs at 400 mm (16 in.) on center filled with loose fill cellulose insulation
Foundation Wall Insulation	Extruded polystyrene 25.4 mm (1 in.)	-
Window Type	Double-pane window	-
Air Conditioner Speed	Single-speed	Single-speed
Heat Recovery Type	None	None
Seasonal Coefficient of Performance (Air Conditioner SEER)	3.81 (SEER 13)	-
Natural Gas Furnace Efficiency	0.80	-
Case Study Building 2	**Full Optimization**	**Optimization with the Significant Variables**
Roofing Material	F12 asphalt shingles	F12 asphalt shingles
Roof Eave Overhang Depth	305 mm (12 in.)	-
Attic Insulation Material	Attic loose fill—R3.3 (IP-R19)	-
External Wall Siding Material	Vinyl siding	Vinyl siding
External Foam Board	12.7 mm (0.5 in.)	-
Wall Core	38 × 89 mm (2 × 4 in.) studs at 400 mm (16 in.) on center filled with loose fill cellulose insulation	38 × 89 mm (2 × 4 in.) studs at 400 mm (16 in.) on center filled with loose fill cellulose insulation
Under-Floor Insulation	Extruded polystyrene 25.4 mm (1 in.)	-
Window Type	Double-pane window	-
Air Conditioner Speed	Single-speed DX	Single-speed DX
Heat Recovery Type	None	None
Seasonal Coefficient of Performance (Air Conditioner SEER)	3.81 (SEER 13)	-
Natural Gas Furnace Efficiency	0.80	-

References

1. Li, Y.; Im, P.; Lee, S.; Bae, Y.; Yoon, Y.; Lee, S. Sensor Incipient Fault Impacts on Building Energy Performance: A Case Study on a Multi-Zone Commercial Building. *Buildings* **2023**, *13*, 520. [CrossRef]
2. Huang, H.; Wang, H.; Hu, Y.-J.; Li, C.; Wang, X. The development trends of existing building energy conservation and emission reduction—A comprehensive review. *Energy Rep.* **2022**, *8*, 13170–13188. [CrossRef]
3. EIA. *March 2015 Monthly Energy Review*; U.S. Energy Information Administration: Washington, DC, USA, 2015.
4. Ruparathna, R.; Hewage, K.; Sadiq, R. Improving the energy efficiency of the existing building stock: A critical review of commercial and institutional buildings. *Renew. Sustain. Energy Rev.* **2016**, *53*, 1032–1045. [CrossRef]
5. Crawley, D.B.; Lawrie, L.K.; Winkelmann, F.C.; Buhl, W.F.; Huang, Y.J.; Pedersen, C.O.; Strand, R.K.; Liesen, R.J.; Fisher, D.E.; Witte, M.J.; et al. EnergyPlus: Creating a new-generation building energy simulation program. *Energy Build.* **2001**, *33*, 319–331. [CrossRef]
6. Klein, S.A.; Beckman, W.A.; Mitchell, J.W.; Duffie, J.A.; Duffie, N.A.; Freeman, T.L.; Mitchell, J.C.; Braun, J.E.; Evans, B.L.; Kummer, J.P.; et al. *TRNSYS 16—A TRaNsient System Simulation Program, User Manual*; Solar Energy Laboratory, University of Wisconsin-Madison: Madison, WI, USA, 2004.
7. Magnier, L.; Haghighat, F. Multiobjective optimization of building design using TRNSYS simulations, genetic algorithm, and Artificial Neural Network. *J. Affect. Disord.* **2010**, *45*, 739–746. [CrossRef]
8. Gossard, D.; Lartigue, B.; Thellier, F. Multi-objective optimization of a building envelope for thermal performance using genetic algorithms and artificial neural network. *Energy Build.* **2013**, *67*, 253–260. [CrossRef]
9. Ghiaus, C. Experimental estimation of building energy performance by robust regression. *Energy Build.* **2006**, *38*, 582–587. [CrossRef]
10. Catalina, T.; Virgone, J.; Blanco, E. Development and validation of regression models to predict monthly heating demand for residential buildings. *Energy Build.* **2008**, *40*, 1825–1832. [CrossRef]
11. Hygh, J.S.; DeCarolis, J.F.; Hill, D.B.; Ranjithan, S.R. Multivariate regression as an energy assessment tool in early building design. *J. Affect. Disord.* **2012**, *57*, 165–175. [CrossRef]
12. Krarti, M. *Energy Audit of Building Systems: An Engineering Approach*; CRC Press: Boca Raton, FL, USA, 2020.
13. Eisenhower, B.; O'neill, Z.; Narayanan, S.; Fonoberov, V.A.; Mezić, I. A methodology for meta-model based optimization in building energy models. *Energy Build.* **2012**, *47*, 292–301. [CrossRef]
14. Bettonvil, B.; Kleijnen, J.P. Searching for important factors in simulation models with many factors: Sequential bifurcation. *Eur. J. Oper. Res.* **1997**, *96*, 180–194. [CrossRef]
15. Rahni, N.; Ramdani, N.; Candau, Y.; Dalicieux, P. Application of group screening to dynamic building energy simulation models. *J. Stat. Comput. Simul.* **1997**, *57*, 285–304. [CrossRef]
16. Corrado, V.; Mechri, H.E. Uncertainty and Sensitivity Analysis for Building Energy Rating. *J. Build. Phys.* **2009**, *33*, 125–156. [CrossRef]
17. Brohus, H.; Heiselberg, P.; Simonsen, A.; Sørensen, K.C. Uncertainty of energy consumption assessment of domestic buildings. In *E-Nova Internationaler Kongress 2009: Null Emissions Gebäude: Gebäude, Gebäudeteile, Anlagen, Anlagenkomponenten, Energie-und Gebäudekonzepte, Proceedings of the E-Nova Internationaler Kongress, Pinkafeld, Austria, 19–20 November 2009*; Fachhochschulstudiengänge Burgenland: Eisenstadt, Austria, 2009; pp. 31–38.
18. Gunay, H.B.; Ouf, M.; Newsham, G.; O'Brien, W. Sensitivity analysis and optimization of building operations. *Energy Build.* **2019**, *199*, 164–175. [CrossRef]
19. Cheng, L.; Bae, Y.; Horton, W.T. A system-level approach for designing multi-family sustainable and energy-efficient housing communities. *Sustain. Cities Soc.* **2019**, *44*, 183–194. [CrossRef]
20. Østergård, T.; Jensen, R.L.; Maagaard, S.E. Early Building Design: Informed decision-making by exploring multidimensional design space using sensitivity analysis. *Energy Build.* **2017**, *142*, 8–22. [CrossRef]
21. Saltelli, A.; Tarantola, S.; Campolongo, F.; Ratto, M. *Sensitivity Analysis in Practice: A Guide to Assessing Scientific Models*; Wiley: New York, NY, USA, 2004; Volume 1.
22. Bro, R.; Smilde, A.K. Principal component analysis. *Anal. Methods* **2014**, *6*, 2812–2831. [CrossRef]
23. Mehmood, T.; Liland, K.H.; Snipen, L.; Sæbø, S. A review of variable selection methods in partial least squares regression. *Chemom. Intell. Lab. Syst.* **2012**, *118*, 62–69. [CrossRef]
24. Bae, Y.; Horton, W.T. An Efficient and Accurate Building Optimization Strategy Using Singular Value Decomposition. In Proceedings of the International High Performance Buildings Conference, West Lafayette, IN, USA, 11–14 July 2016; Paper 199.
25. Bae, Y.J. Integrated Design Tool of Building System Optimization for Building Life Cycle Cost. Ph.D. Thesis, Purdue University, West Lafayette, IN, USA, 2016.
26. Kennedy, J.; Eberhart, R.C. A discrete binary version of the particle swarm algorithm. In Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics. Computational Cybernetics and Simulation, Orlando, FL, USA, 12–15 October 1997; Volume 5, pp. 4104–4108.
27. Kennedy, J. Swarm intelligence. In *Handbook of Nature-Inspired and Innovative Computing: Integrating Classical Models with Emerging Technologies*; Springer: Boston, MA, USA, 2006; pp. 187–219.
28. PNNL Residential Prototype Building Models. Available online: https://www.energycodes.gov/development/residential/iecc_models (accessed on 30 July 2013).

29. RSMeans. RSMeans Residential Cost Data 2013. Available online: http://www.rsmeans.com (accessed on 30 July 2013).
30. Rushing, A.S.; Kneifel, J.D.; Lippiatt, B.L. *Energy Price Indices and Discount Factors for Life-Cycle Cost Analysis-2013: Annual Supplement to NIST Handbook 135 and NBS Special Publication 709*; National Institute of Standards and Technology: Gaithersburg, MD, USA, 2013.
31. Parsopoulos, K.E.; Vrahatis, M.N. Recent approaches to global optimization problems through Particle Swarm Optimization. *Nat. Comput.* **2002**, *1*, 235–306. [CrossRef]
32. Ljung, L.; Söderström, T. *Theory and Practice of Recursive Identification*; MIT Press: Cambridge, MA, USA, 1983.
33. Goodwin, G.C.; Sin, K.S. *Adaptive Filtering Prediction and Control*; Courier Corporation: Chelmsford, MA, USA, 2014.