A DESIGN AND ANALYSIS OF COMPUTER EXPERIMENTS APPROACH FOR GREEN BUILDING

by

MARJAN SAYADI

Presented to the Faculty of the Graduate School of

The University of Texas at Arlington in Partial Fulfillment

of the Requirements

for the Degree of

MASTER OF SCIENCE IN INDUSTRIAL ENGINEERING

THE UNIVERSITY OF TEXAS AT ARLINGTON

December 2016

Acknowledgements

My background in my undergraduate studies was in Statistics. Although I was always interested in Statistics, I never considered pursuing my education in the same field until I took the Applied Regression Analysis course in the first year of study for my Master's of Science degree program in the Department of Industrial, Manufacturing, & Systems Engineering (IMSE) at the University of Texas at Arlington. The instructor of the course for that specific semester was Dr. Victoria Chen, and during the semester, I learnt how much real world problems can be mixed with science, math and statistics to name a few. Thus, I decided to implement my knowledge in a real world problem by changing my Master's of Science from the non-thesis program to the thesis-based program, although it was much easier to graduate by taking some theoretical courses. I needed to have a supervisor to guide me through the research; although there are a lot of knowledgeable professors in the IMSE Department, I decided to work with Dr. Chen. Here, I would like to express my gratitude and thankfulness for her patience and support during my Master's program, and specifically during my thesis.

I would like to appreciate Dr. Jay M. Rosenberger for all his guidance and help during my thesis studies. I had a chance to be his student in the Operation Research course, during which I learned a lot from his teaching methods and knowledge. I am also so grateful to my committee member Dr. Shouyi Wang who supported me during my thesis studies.

I was also blessed to have hard working classmates, Naveen Kumar Thiagarajan and Rahul Ramakrishnan Ramesh, helping me. They supported me by providing me with some of the experimental runs in eQUEST and ATHENA and saved me a great amount of time. I also need to thank Yasaman Behain and Shirish Rao who helped me with the statistical analysis using multivariate adaptive regression splines. Many thanks to my supportive friends in the Center on Stochastic Modeling, Optimization, & Statistics (COSMOS) who shared their knowledge and always encouraged me not to give up.

I have been so blessed to have Farshad Zahedi as my husband and as a big supporter in my academic and personal life. He is always my strongest motivator to be ambitious. He motivated me to start my Master's of Science degree program and supported me emotionally during my studies at UTA.

Finally, I am not able to say in word how much I am grateful to have my family beside me. Although they are not beside me physically, I always have their love and support with me. My lovely parents, Shahin Hashemi and Mohammad Sayadi, and my supportive siblings Maryam and Nader.

November 9, 2016

Abstract

A DESIGN AND ANALYSIS OF COMPUTER EXPERIMENTS APPROACH FOR GREEN BUILDING

Marjan Sayadi, MS

The University of Texas at Arlington, 2016

Supervising Professor: Victoria C. P. Chen

The coming shortage of energy sources and critical environmental impacts are two major factors that have forced a change in product design processes. The shortage of energy sources is related to the limitation on non-renewable energy sources on earth and requires the development of new concepts with lower energy consumptions. Environmental impact, on the other hand, is concerned about the negative effects of products on the natural environment. In recent years, research on designing more environmentally-friendly products that consume lower of energy with lower environmental impact has been initiated to address these issues [1-3]. Designing new powertrains for vehicles [4] and conducting research on developing airplanes with new sources of energy [5] are some of the examples. Building structures are of great interest, since the building have a significant impact on environment and energy consumption [6].

Buildings can be designed so that their energy consumption is reduced, by using new materials with higher thermal resistance, or implementing new design strategies to reduce the heat extraction from the building. In addition, there is a certain life cycle for any structure, which includes the time span between the manufacturing of the materials to the decomposition of these materials, and is called “cradle-to-grave” [6]. This cycle is usually

used as a criterion for the environmental effect, and minimizing this effect is of great interest.

While it is desirable to simultaneously minimize both energy consumption and environmental impacts, it is not straightforward to achieve because these two objectives depend on variety of factors. Therefore, it would be helpful to implement a multiple-objective optimization approach to design a building that satisfies both objectives. Buildings are complicated structures that include different subsystems, making it a multivariate, multi-response case study.

In this study, two computer experiments are designed to evaluate the performance of a building with the focus on the energy consumption and the environmental impact. Since building variables include both categorical and continuous variables, two different design of experiments approaches are used to combine them together. The computer simulation of the energy consumption is performed in eQUEST [7], while the environmental impact is calculated in ATHENA impact estimator software [8].

The goal of the current work is to compare different experimental designs and different statistical modeling methods to help inform our approach for a multivariate, multi-response framework. For this purpose, a residential building is considered as the case study, and different design factors that affect energy consumption and the environmental impacts of the building are identified. A design of experiment is implemented to realize the simulations that can provide the data to study both performance objectives. Finally, the results of the experiments are studied using treed regression and multivariate adaptive regression splines approaches to identify important factors.

Table of Contents

List of Illustrations

List of Tables

CHAPTER 1

INTRODUCTION

1.1 Background

Energy efficient, environmentally-friendly buildings are called “green buildings,” and in order to achieve such a design, it is required to simulate the performance of the building prior to construction. While a building simulation tool can allow exploration of many designs, a comprehensive exploration requires an organized approach instead of “trial and error.” The goal is to reach to a building design that can simultaneously achieve low energy consumption and low environmental impact.

A comprehensive exploration of building structures requires a lot of factors, which can be categorized as follows:

- Electrical subsystems
- Wells and septic system
- Wall system
- Building orientation and footprint
- Foundation system
- Door system
- Window system
- Roof system
- Plumbing system
- Ventilation system
- Heating and cooling system
- Landscaping system

In addition, the building simulation output includes different aspects, leading to a multiple performance metrics to consider. Some of the outputs are as follows:

- Energy consumption
- Environmental impact
- Life cycle cost analysis

The ultimate goal would be to optimize the building options to achieve a green building design that simultaneously considers all the performance outputs mentioned above. To conduct the optimization, a multi-objective approach is needed to handle the multiple performance outputs. In order to calculate the performance outputs, implementation of different building software tools are needed. Unfortunately, the menu-driven format of building software tools makes them difficult and cumbersome to implement within an optimization routine. Hence, a surrogate optimization approach (e.g., [9]) is needed. The surrogate optimization approach for our problem will construct metamodels of the performance outputs using a design and analysis of computer experiments (DACE) approach [10, 11]. The resulting metamodels can then be employed within an optimization routine to represent the multiple performance objectives. The focus of this thesis is to study two different methods for creating experimental designs for the DACE approach.

In this thesis, the primary challenge for the surrogate optimization approach is handling a mix of many discrete and continuous input variables, where discrete variables include both categorical factors and discrete-numerical variables. Classical experimental design [12, 13] is appropriate for categorical factors, where a continuous variable can be converted into categories by partitioning the continuous range into discrete subranges. In DACE, the factor variable space is commonly assumed to be continuous. In this case, space-filling experimental designs are preferred. A pseudo-random generation of points in a space qualifies as a space-filling design, but more uniformly-spaced designs can be achieved via factorial designs or quasi-random designs. In this thesis, the experimental designs studied were created to handle the mix of discrete and continuous input variables.

The first experimental design was introduced by Kung (2012) [14]. Kung's design utilizes a classical experimental design, specifically a mixed orthogonal array [15], for the discrete input variables and a quasi-random experimental design, specifically a Sobol' low-discrepancy sequence [16], for the continuous input variables. The challenge then lies in merging the mixed array and the Sobol' sequence. Kung (2012) employs a Latin hypercube design [17] to achieve this merge. The second experimental design was described by Martinez (2013) [18].Martinez starts with a quasi-random experimental design, specifically a Sobol' sequence, then conducts a "rounding" method to convert continuous values to a set a discrete values or categories for the discrete input variables. Finally, in order to construct metamodels, treed regression [19, 20] and multivariate adaptive regression splines (MARS) [21] methods are used, and results are compared. The green building performance outputs from two software tools are studied. For energy consumption, eQUEST software [22] is used, and for environmental impacts ATHENA impact estimator for buildings [8] is used.

1.2 Organization

This thesis includes six chapters. Chapter 2 describes background of the research, including green building software tools and a review of the literature. The computer model setup is provided in Chapter 3, while the design of experiments and statistical analysis methods of the first and the second designs are introduced in Chapters 4 and 5, respectively. The discussion and validation of results are provided in Chapter 6. Finally, conclusions are discussed in Chapter 7, followed by suggestions for the future research.

CHAPTER 2

LITERATURE REVIEW

2.1 Green Building Simulation Tools

2.1.1 eQUEST

One of the commonly used tools to simulate energy usage in buildings is the QUick Energy Simulation Tool (eQUEST) [22]. eQUEST was initially developed as a part of the Energy Design Resources program, which was administered by Pacific Gas and Electric Company, San Diego Gas & Electric, and Southern California Edison. This software provides inexperienced users with the ability to develop simulation models of buildings, by utilizing a building creation wizard, an energy efficiency measure (EEM) wizard, and graphical reporting [23].

eQUEST accepts properties of buildings, i.e., location, orientation, wall/roof construction, and window properties, and properties of the building subsystems, including heating, ventilation and air conditioning (HVAC) systems, day-lighting and various control strategies, and evaluates the effect of these variables on any single or combination of energy conservation measures. The software is capable of modelling the buildings as simple as a box and single-zone to as complex as the actual design imported from AutoCAD with complex schedules and rate schedules. Inputs to the program are broken into "schematic design" and "design development," and the software calculates the annual energy consumption and associated costs for a particular building design based on the provided inputs. In addition, the software provides an extensive report, which includes a summary of inputs, e.g., schedules, building construction characteristics, a summary of load components and peak loads, characteristics of the HVAC system, including the input characteristics, system size, runtimes, capacity, and air/fluid flow, and hourly reports from user-specified building components. However, eQUEST has some limitations. For example, HVAC system types included in the software are limited to the predefined options [7].

The eQUEST software tool is menu-driven, and parameters used to simulate a building can be defined in the software and in a straightforward approach. A sample screenshot of one of the menu windows of the software is shown in Figure 2.1. The parameters used to generate a building model are completely discussed in the following sections.

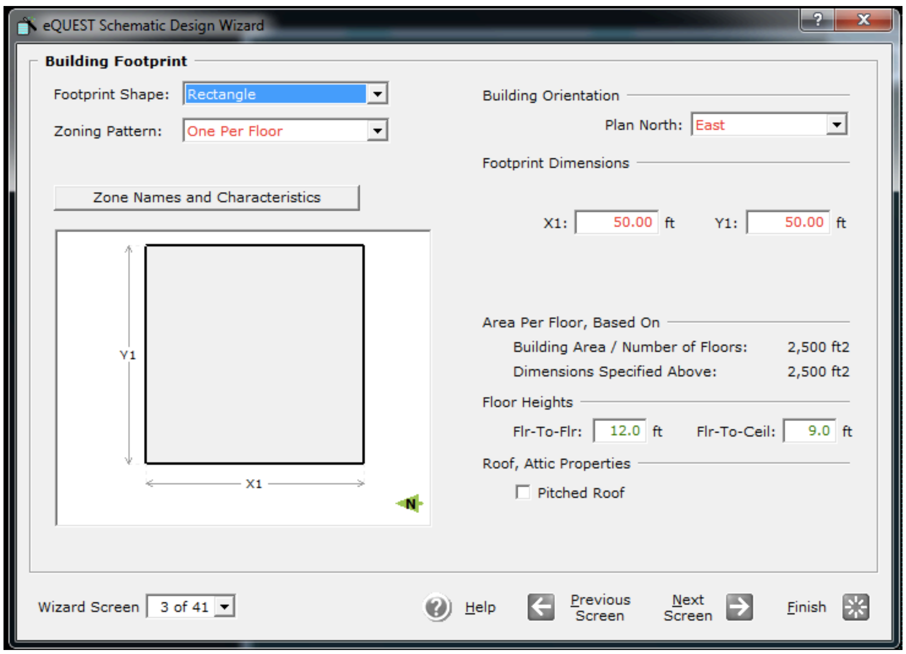


Figure 2.1 A screenshot of the model setup window in eQUEST

Several performance metrics are available in eQUEST, which calculates the type of energy consumption in the building. These performance metrics are as follows:

- Annual source energy (Total in million British thermal unit (Mbtu) and EUI in kBtu)
- Annual energy usage (Electricity in kW and Natural Gas in Therms)
- Lighting (Electricity in kW)

- HVAC energy (Electricity in kWh, Natural Gas in Therms, and Total in Mbtu)
- Peak (Elect in kW and Cooling in Tons)

Only three of these performance metrics are not highly correlated, which are Annual source energy (Total in Mbtu), Annual energy usage (Elect in kW), and HVAC energy (Electric in kW) [14]. Thus, in this study, only these three performance metrics are investigated from eQUEST.

2.1.2 ATHENA Impact Estimator for Building

First released in 2002, ATHENA Impact Estimator [8] is developed to asses and compare the environmental impact of building designs. It is the only open-source free software available in North America that can model and simulate a complete building with the assemblies based on the life cycle assessment (LCA) methodology [8]. The software is equipped with different impact estimation methods, including mid-point impact estimation methods developed by the U.S. Environmental Protection Agency (EPA), reported for the reduction and assessment of chemical and other environmental impacts. The software is able to simulate a wide range of industrial, institutional, commercial, and residential buildings. It can model over 1,200 structural and envelope assembly combinations, design new buildings and major renovations, and distinguish between owner-occupied and rental facilities. The outputs comprise the flows from and to nature: energy and raw material flows plus emissions to air, water and land.

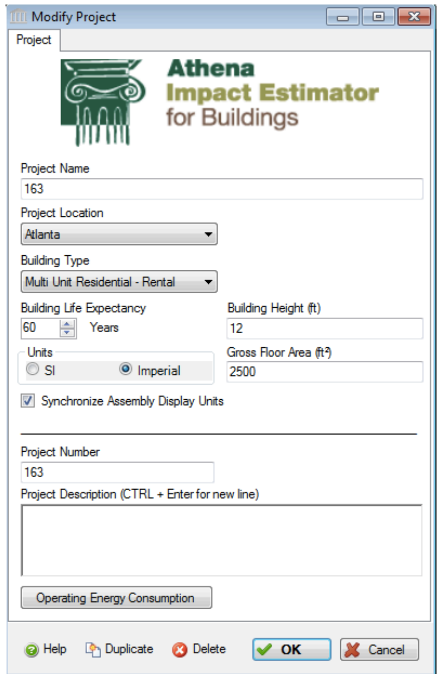


Figure 2.2 A screenshot of the model setup window in ATHENA Impact Estimator software

Some of the analysis capabilities of the software are:

- Global warming potential (GWP) in kg CO2 eq
- Acidification potential in kg SO2 eq
- HH particulate in kg PM2.5 eq
- Eutrophication potential in kg N eq
- Ozone depletion potential in kg CFC-11 eq
- Smog potential in kg O3 eq
- Total primary energy in Mega Joule (MJ)
- Non-renewable energy in MJ
- Fossil fuel consumption in MJ

ATHENA is also menu driven, and setting up the parameters of the building is straightforward. A sample screenshot of the software is shown in Figure 2.2, and the parameters used to generate the building are completely discussed in the following chapters.

2.1.3 SPM (Salford Predictive Modeler)

The SPM (Salford Predictive Modeler) software suite [24] is an integrated suite of data mining software. It includes CART, MARS, TreeNet, and Random Forests modules, among which only CART and MART modules are used in this research, which are briefly introduced.

CART (Classification and Regression Trees) is a robust decision-tree tool for data mining, predictive modeling, and data preprocessing. CART trees can be used to generate accurate and reliable predictive models for a broad range of applications from bioinformatics to risk management and new applications are being reported daily. Salford Systems' CART is based on the original CART code developed by Stanford University and University of California at Berkeley statisticians Breiman, Friedman, Olshen and Stone [20]. The CART module accepts the training and testing datasets in delimited format, and the parameters used to create the trees are set in the “Model Setup” window. The “Model” tab is used to import and select the predictor and the target variables, while the “Testing” tab is used to import the testing dataset. The “Limit” window is used to select the appropriate parameters for tree generation. A screenshot of the Model window is shown in Figure 2.3.

Multivariate Adaptive Regression Splines (MARS) has become widely known in the data mining and business intelligence. MARS is a flexible and automated regression modeling tool that automates the building of accurate predictive models for continuous and binary dependent variables.

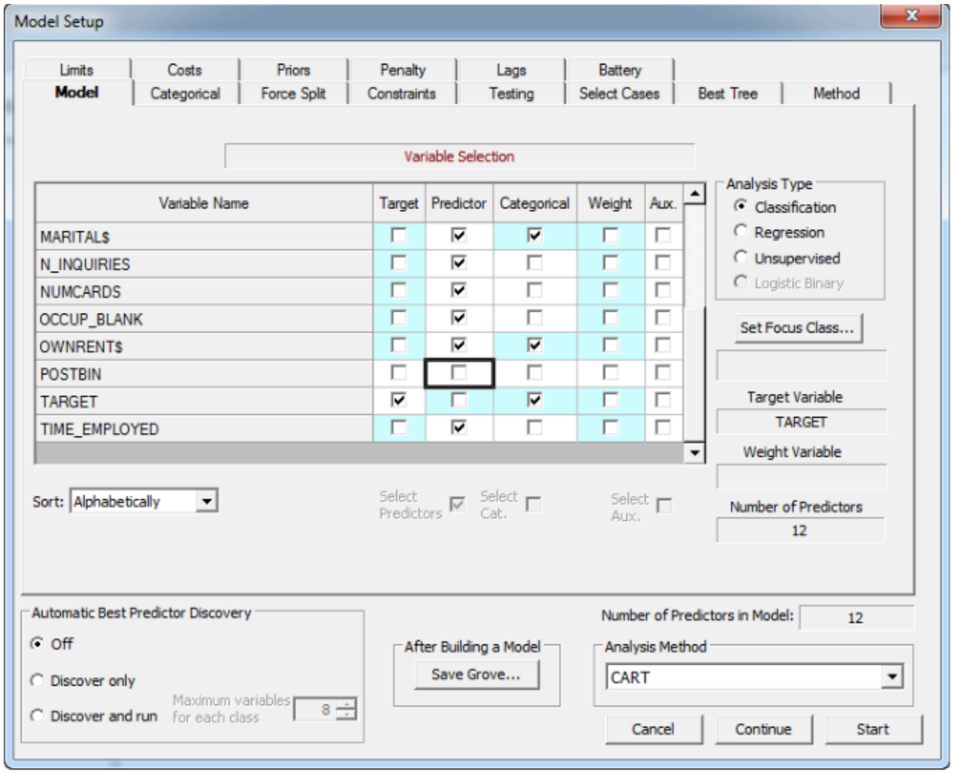


Figure 2.3 A screenshot of the model setup window in SPM Salford Systems software

Similar to the CART module, the MARS module accepts the training and the testing datasets in a comma separated variables (CSV) format. After importing, the rest of the modelling is done in the model setup window. For example, since the response variable is continuous, regression should be chosen as the analysis type. All of the input variables are shown in the "Variable Selection" section in the "Model" tab, where the response and target variables can be defined. The analysis method is chosen in the lower right corner of the window. Both, CART and MARS modules, are designed in a menu-driven format, and the parameters can be easily set in these menus. The parameters used to design the models based on the MARS method are completely discussed in Chapter 5.

2.2 Green Building Optimization Research

Approximately one third of our primary energy supply is consumed in buildings, and consequently, buildings are a primary contributor to global warming and ozone depletion [25]. In addition, at the global level, civil works and building construction consumes 60% of the raw materials extracted from the lithosphere, and from this volume, building represents 40%, in other words 24% of these global extractions [26]. These facts enforce engineers to design structures that consume less energy and have lower environmental impact. Green building is a recent design philosophy that requires the consideration of resources depletion and waste emissions during its whole life cycle [27]. A green building is designed with strategies that conserve resources, reduce waste, minimize the life cycle costs, and create healthy environment for people to live and work [28]. Green buildings are a promising design for the future urban settings. Designing a building with specifications of a green building includes two major parts.

One major part of design is reducing the energy consumption of the building by using more thermal resistant materials and involving the effect of the building schematic on the energy dissipation. The concept of green building, when it refers to the building components, have been in research for a long time. Wong et al. [2], for example, considered the roof top garden as a part of a green commercial building, and showed that a 0.6-14.5% reduction in the energy consumption is seen with the roof top garden. In a more general point of view, the effect of using renewable energies has also been studied. Solar energy, as a very important source, has been studied as an energy source in green buildings, and it was shown that after one year of operation, the solar system was found to contribute 70% of total energy [29].

To achieve the goals of sustainability it is required to adopt a multi-disciplinary approach covering a number of features such as: energy saving, improved use of materials, reuse and recycling and emissions control [30], and, therefore, design of experiments is used to obtain the optimal building design.

2.3 Design of Experiment

Since in this study two menu-driven software tools are used and many factors are identified to be modeled as the building factors, it will be very time-consuming to simulate all of the possible cases. Thus, design of experiments (DOE) is used to save time and effort [31]. DOE is a systematic method to study the relationship between factors affecting a process and the output of that process, i.e., cause-and-effect relationships. DOE is used to systematically select a limited number of experiments from a large number of possible experiments. The DOE methods used in this research are explained in this section.

2.3.1 Kung's design

In 2012, Pin Kung who was a Ph.D. student in the University of Texas at Arlington, was the first to study a DACE framework for eQUEST [14]. The building factors were categorized into twelve main categories, shown in Table 2.1, and the availability of each of these factors in eQUEST and ATHENA are noted in the parentheses in Table 2.1. As mentioned, a design of experiment (DOE) was developed [14] to limit the number of experiments. Since, there are two types of variables, two different methods are used to design the experiments for either of the discrete and continuous variables separately. Then, another method is used to combine these two design into a single design.

The experimental design of the discrete variables was performed using a mixed orthogonal array (MA) [15]. Classical orthogonal arrays require the same number of levels for each factor dimension. Mixed arrays allow factors with different numbers of levels. Continuous variables were handled using the Sobol' sequence. Sobol sequences are an example of quasi-random low-discrepancy sequences [32]. In this research, the MATLAB function sobolset is implemented to create the design of experiments using the Sobol' sequence.

Finally, a two-factor Latin hypercube method [14] was adapted, to combine the discrete and continuous variables in a single design. One factor of the Latin Hypercube selects one row from the mixed array output, and the other factor selects one row from the Sobol' sequence, and these two rows are concatenated to create one row of the combination of these experiments. Latin hypercube sampling is a statistical method for generating a near-random sample of parameter values from a multidimensional distribution [33], and is often used to construct computer experiments. The Latin hypercube design used by Kung (2012) [14] has a frequency parameter of 2, which means that each row of the mixed array and each row of the Sobol' design appears twice in the combined design. This experimental design is used as the training data set in this thesis. Kung's design used to generate the testing dataset is generated using a MATLAB code [14].

Table 2-1 Stages and Decision Variables for Green Building [14]

Stage	Building Stage with Options
1	Siting Options • Orientation and Footprint (eQUEST)
2	Electrical System • AC System (eQUEST) • Both AC and Solar System • Solar System
3	Wells and Septic System • Concrete Septic Tank • Fiberglass Septic Tank
4	Foundation System • Concrete Ground Floor (eQUEST) • Concrete Slab on Grade (ATHENA) • Generic Portland Cement • Steel Foundation System
5	Plumbing System

	• Freshwater System • Greywater System • Rainwater Catchment System
6	Wall System • Concrete Wall (ATHENA, eQUEST) • Curtain Wall (ATHENA) • Drywall • Metal Frame (eQUEST) • Straw Bale Walls • Wood Frame (eQUEST)
7	Window System • Clear/Tint Windows (eQUEST) • Glazed Windows • Low-e Windows (eQUEST) • Reflective Windows (eQUEST) • Wood Frame Windows (ATHENA, eQUEST)
8	Door System • Steel Door (ATHENA, eQUEST) • Wood Door (eQUEST)
9	Roof System • Concrete Tile Roof (ATHENA, eQUEST) • Generic Fiber Cement Roof • Roof Surface Materials (eQUEST)
10	Ventilation System • Balanced Ventilation System • Exhaust Ventilation System • Supply Ventilation System • Ventilation-Activity Areas (eQUEST)
11	Heating and Cooling System • Fan System (eQUEST) • HVAC System (eQUEST)
12	Landscaping System • Sprinkler System

2.3.2 *Martinez's Design*

Nadia Martinez, in 2013, studied the global optimization of nonconvex piecewise linear regression splines [18]. In her research, a method was generated for design of experiments to handle a mix of variable types. Specifically, Sobol' sequence is first generated with a range of 0 and 1. Then for discrete variables, the following method is applied: a 2-level variable takes the first level if the relative value in the Sobol' sequence is less than 0.5, otherwise it takes the second level. For variables of levels higher than 2, the threshold is calculated using the following method. In addition, as discussed in section 2.3.1, for the fourteen variables that are considered as continuous variables, a 96-by-14 matrix was generated using Sobol' sequence, and each column of the matrix was randomly assigned to one of the continuous variables. Since Sobol' sequence output is between zero and one, the number generated by Sobol' sequence is scaled to the relative range of each continuous variable, and then use the generated numbers as the value of the respective continuous variable. For example, for a variable with 4 levels, the τ value is 0.62996by using the following formula

$$\tau = \frac{1}{\sqrt[p-1]{p}}$$

where p represents the number of levels. Thus, if the maximum value of all of the relative values for the variable for each level in Sobol' sequence is equal or greater than 0.62996, the variable takes the level that corresponds to the maximum value, otherwise it takes the last level. For example, assume that for the first run $X_{2,}$ which is the discrete variable with four levels, takes 0.23375 regarding the scaled value in continuous space for the first level, 0.73888 for the second level, and 0.14667 for the third one. The maximum value of these three numbers is 0.73888, which corresponds to the second level and is greater than

0.62995 (the threshold for 4-level variables). Therefore, in the first experimental run, X_2 should take its second level, which is “Face South.”

2.4 Statistical Modeling

2.4.1 Treed regression

Classification and regression trees are machine-learning methods for constructing prediction models from data [34]. The models are obtained by recursively partitioning the data space and fitting a simple prediction model within each partition. As a result, the partitioning can be represented graphically as a decision tree. Classification trees are designed for dependent variables that take a finite number of unordered values, with prediction error measured in terms of misclassification cost. Regression trees are for dependent variables that take continuous or ordered discrete values, with prediction error typically measured by the squared difference between the observed and predicted values. It is widely used to handle categorical and continuous variables [20, 21]. In order to apply the treed regression model, at first it is required to determine the trees. The CART module of the SPM [24], which is powerful tool widely used for data mining, is adopted to generate the trees. The parameters used to generate the regression trees are completely discussed in Chapter 5. After tree generation part, it is needed to fit the regression line on terminal nodes (TNs). This is done by using the following equation:

$$\hat{g}_{tree} = \sum_{j=1}^{J} \hat{Y} \,.\, I\{X \in R_j\}$$

where J is the number of TN’s, I is an indicator function, R’s are the disjoint regions (tree TNs), and $\hat{Y}$ is defined as follows:

$$\hat{Y} = \beta_0 + \sum_{j=1}^{J} \beta_j X$$

where β is the vector of the coefficients in the regression model and J is the number of TNs. In this research, Statistical Analysis Software (SAS) [35] is used to fit the linear regression models using the data partitioned at the TNs.

2.4.2 Multivariate Adaptive Regression Splines

In 1991, Friedman [21] proposed multivariate adaptive regression splines (MARS) to fit a piecewise regression model to a multivariate set of predictor variables based on a single response variable. The MARS model is defined as follows:

$$\hat{f} = C_0 + \sum_{i=1}^{k} C_i B_i(x)$$

where C_i is a constant coefficient, $B_i(x)$ is the basis function (BF), and x is a vector of predictor variables. The basis function is defined as $\max(0, constant - x)$ or $\max(0, x - constant)$ for the numerical and continuous variables, and for categorical variables an indicator function same as treed regression model is used, where the R's are the regions in which X is valid. In this research, the MARS module from Salford Systems [24] is used to apply the MARS approach.

CHAPTER 3

COMPUTER MODEL SETUP

The purpose of this study is to provide a comprehensive green building decision-making framework to be used in the DACE approach. Building factors were divided into twelve main categories [14], shown in Table 2.1. Two different software packages, eQUEST and ATHENA, were used to simulate the building performance metrics, and therefore, the available levels of these performance metrics is different, and is specified in parentheses in Table 2.1.

Some of these options are only available in one of the tools. For example, in stage 1 (orientation siting), only eQUEST has an option for the “Orientation and Footprint..” Some of these stages, on the other hand, cannot be modeled in either eQUEST or ATHENA. For example, for stage 12, “sprinkler system” option cannot be modeled in either of them. After reviewing the availability of these factors in eQUEST and ATHENA, 52 factors were found to be available in eQUEST or in ATHENA, or in both. For some input factors, the software tool only permitted a specific and finite set of levels. Hence, these are discrete (finite) factors. A factor is considered as continuous if there is no specific option limiting its value in the software. Therefore, among these 52 decision variables, 38 variables are discrete and 14 variables are considered continuous. The values, however, must be selected from a specific default range defined in eQUEST (see Table 3.4).

The factors available in both software tools were explored to find the common options across both tools. Since it is important to have similar options for similar factors in eQUEST and ATHENA as much as possible, the options for the 38 discrete variables are considered as shown in Table 3.1. In addition, as it is explained later in the Section 4.1, the options for each factor are limited to 23 variables with two options, one variable with three options, and fourteen variables with four options.

Table 3-1 38 discrete decision variables

Stage	Building category	Variable	38 discrete variables	Options in eQUEST or ATHENA
1	siting options	Foot Print Shape-X1	1) 25*100 ft2 2) 50*50 ft2	option in both
2	Electrical System	Based on Default from eQUEST and No option in ATHENA		
3	Well and Septic System	No option in eQUEST and ATHENA		
4	foundation system	Orientation-X2	1) N/S Component (Face North) 2) N/S Component (Face South) 3) E/W Component (Face East) 4) E/W Component (Face West)	no option in ATHENA
		Ground Floor Interior insulation-X3	1) 1 inch Polystyrene expanded (R-4) 2) 1 1/2 inches Polystyrene expanded (R-6) 3) 2 inches Polystyrene expanded (R-8) 4) 3 inches Polystyrene expanded (R-12)	option in both
		Ground Floor Construction-X4	1) 4inch Concrete 2) 8inch Concrete	option in both

		Concrete slab on grade-X5	1) 3000psi 2) 4000psi 3) 5000psi 4) 6000psi	no option in eQUEST
		Ground Floor Exterior/Cav insulation-X6	1) No Batt 2) R-11 Batt(3.5484inch) 3) R-19 Batt(6.1296 inch) 4) R-30 Batt(9.678 inch)	option in both
		Ground Floor Cap-X7	1) 1.25 inch Lightweight Concrete 2) 2 inch Lightweight Concrete 3) 3 inch Lightweight Concrete 4) 4 inch Lightweight Concrete	no option in ATHENA
5	Plumbing System	No option in eQUEST and ATHENA		
6	Wall System	Walls Interior Insulation-X8	1) None 2) R-4 Polystyrene(1 inch)	option in both
		Walls Exterior insulation-X9	1) R-4 Polystyrene (1 inch) 2) R-6 Polystyrene (1 1/2 inches) 3) R-8 Polystyrene (2 inches) 4) R-12 Polystyrene (3 inches)	option in both
		Walls Additional insulation-X10	1) No Mineral Batt 2) R-11 Mineral Batt (3.5484inch)	option in both

		Walls Construction-X11	1) Wood Frame, 2Í4, 16 inch O.C. 2) Wood Frame, 2Í4, 24 inch on center	option in both
		Walls Type-X12	1) None Load Bearing 2) Load Bearing	no option in eQUEST
		Walls Sheathing Type-X13	1) None 2) oriented strand board (OSB)	no option in eQUEST
		Walls External finish-X14	1) Concrete(Brick-Concrete) 2) Brick(Brick, Ontario-standard)	option in both
		Walls External Color-X15	1) Light 2) Dark	no option in ATHENA
		Walls External Color Type-X16	1) Alkyd solvent based 2) latex water based	no option in eQUEST
7	Window System	Window Type-X28	1) Operable 2) fixed	option in both
		Window Frame Type-X29	1) Aluminum clad wood 2) vinyl clad wood	option in both
		Window Spacer Type-X30	1) Mtl 2) Ins	no option in ATHENA
		Number of panes of Window Frame-X31	1) 2 2) 3	no option in eQUEST
		Window Glass category-X32	Low-e:	option in both

			1) Double 2) Triple	
		Window Glass type-X33	1) Clear 1/8, ¼ inch 2) Clear 1/8, ½ inch 3) Clear ¼, ¼ inch 4) Clear ¼, ½ inch	no option in ATHENA
		Window Glass Glazing-X34	1) Soft 2) Hard	no option in eQUEST
		Total Window Area % North-X35	1) 6% 2) 12% 3) 18% 4) 24%	option in both
		Total Window Area % South-X36	1) 6% 2) 12% 3) 18% 4) 24%	option in both
		Total Window Area % East-X37	1) 6% 2) 12% 3) 18% 4) 24%	option in both
		Total Window Area % West-X38	1) 6% 2) 12%	option in both

			3) 18% 4) 24%	
8	Door System	Door Construction(fixed)	opaque, Wood solid core Flush1-3/8 in	option in both
		Door Dimensions-Height& Width(fixed)	7 feet and 2.67 feet	option in both
9	Roof system	Roof Construction-X17	1) Wood Advanced Frame, 24 inch on center 2) Wood Advanced Frame, >24 inch on center	no option in ATHENA
		Roof Load Bearing-X18	1) 50 psi (Pound Per Square Inch) 2) 100psi	no option in eQUEST
		Roof Exterior insulation-X19	1) None 2) R-8 Polystyrene Expanded (2 inches) 3) R-20 Polystyrene Expanded (4 inches) 4) R-30 Polystyrene Expanded (6 inches)	option in both
		Roof Additional insulation-X20	1) no Batt and no barrier 2) R-11 Mineral Batt (3.5484inch) 3) R-19 Mineral Batt (6.1296 inch) 4) R-30 Mineral Batt (9.678 inch)	option in both
		Ceiling Batt insulation-X21	1) R-11 Batt 2) R-13 Batt 3) R-19 Batt	no option in ATHENA

			4) R-30 Batt	
		Ceiling interior finish-X22	1) Drywall Finish 2) Plaster Finish	no option in ATHENA
		Ceiling Exterior finish-X23	1) Concrete 2) Clay tile 3) Asphalt pavement, weathered	option in both
		Ceiling Color-X24	1) Light 2) Dark	no option in ATHENA
		Roof Type-X25	1) Without Pitched (Parallel) 2) With Pitched	option in both
		Roof Decking type-X26	1) Plywood 2) OSB	no option in eQUEST
		Roof Decking thickness-X27	1) ½ in 2) ¾ in	no option in eQUEST

In this study, a low rise residential building located in Atlanta, GA is assumed as the case study. The building type is assumed to be a single-family rental building. Some of the building characteristics are assumed to be fixed in both software tools. The area of the building is 2500 square-feet with a 60-year life expectancy, and the zoning pattern is one per floor. The building is assumed over crawl space and the ground floor finish is carpet (no pad). The roof dimension is assumed 25-by-100 or 50-by-50 feet, based on the foot print size, and no insulation is assumed for the top floor ceiling rigid (below the attic) and the framing is assumed as wood-standard framing. Two shapes of footprint are assumed, a rectangular footprint shape with width and length of 25 and 100 feet, and a square footprint with a width and length of 50 feet. It is assumed for the rectangular footprint to have a 100*12 square-feet exterior wall at north and south, and a 25*12 square-feet exterior wall at east and west of the building. In addition, for the square shape footprint, all four exterior walls are assumed with a size of 50*12 square-feet. The following assumptions are also considered for windows: the window width and height are 3 and 6 feet, respectively. The window sill height is 2 feet, and the window frame width is 1.3 inches. Since the window area is $3 \times 6 = 18$ square-feet and fixed, the window area percentage depends on the three available options for the wall size, 50×12 and 100×12 and 25×12 square-feet. Therefore, the number of windows allocated to each wall is computed from the areas of the window and wall using the following relation:

$$number\ of\ windows\ on\ each\ wall = \frac{window\ area\ \% \times wall\ area}{window\ area}$$

Table 3.2 shows the options for the window area% and the number of windows based on the wall size. For example, when the wall size is 100×12 square-feet, and the window is 6% of the wall area, the number of the windows is calculated by multiplying 6% by 1200 and dividing it by 18, which yields 4 for the number of windows. In addition, the “low-e”

Option (e2=0.1) is considered as the window glass type while air is considered as the window glass coating type.

Two doors are considered, one front door and one back door. The position of these doors depends on the orientation of the building. If it is facing north or south, one door is located in the north wall and the other one is located in the south wall. Otherwise, the doors are located in the east and west walls, one per wall. The width and height of the doors are assumed to be 2.67 and 7 feet, respectively. The doors category is assumed to be the wood solid core flush with 3/8inch width and the door type is opaque.

Seven different types of area activity are considered in eQUEST, and the percentage of the area or each type is shown in Table 3.3 [14]. These numbers are based on [14] with a few changes. Since in this study a garage is not included with the building, this option has not been considered. The rest of the activity areas were also modified in percentage in order to have a 100% total activity. There is a default range for the maximum number of occupants in square-feet and the design ventilation in cubic-feet per minute (CFM). Thus, 14 variables are considered as continuous variables. Table 3.4 represents these fourteen continuous decision variables and their ranges. The remaining factors in each software tool, eQUEST or ATHENA, are assumed as default values, since these variables are specific to each software and are not shared between them.

Table 3-2 Number of the windows based on the wall size and window area%

window % area	wall size		
	100*12	25*12	50*12
	number of window		
6%	4	1	2
12%	8	2	4
18%	12	3	6
24%	16	4	8

Table 3-3 Table of Percent Area of Residential Low Rise ([14] with some adjustments)

Activity Area Type	Detailed Items	% Area
1. General Living Space	Family/Den (400) + Living Room (400) + Bath#1 (40) + Bath#2 (40) + Bath-Master (70) + Closets (125) = 1075	43%
2. Bedroom	Bedroom#1 (180) + Bedroom#2 (180) + Bedroom#3 (138) + Bed-Master (252) =750	30%
3. Dining Area Dining Room	=250	10%
4. Kitchen and Food Preparation	Kitchen (109) + Pantry (16) + Breakfast (50) =175	7%
5. Corridor Hall	Hall = 75	3%
6. Laundry	Laundry = 50	2%
7. All Others Entry	= 125	5%
Total: 2500 Square Feet		

Table 3-4 Fourteen Continuous decision variables ([14] with some adjustments)

VARIABLE	**Description**
Max occupancy – bedroom-X39	Range: 575 to 675
Ventilation – bedroom-X40	Range: 10 to 30
Max Occupancy – living space-X41	Range: 575 to 675
Ventilation – living space-X42	Range: 10 to 30
Max occupancy – Dining area-X43	Range: 5 to 105
Ventilation – Dining area-X44	Range: 10 to 30
Max occupancy – kitchen-X45	Range: 250 to 350
Ventilation – Kitchen-X46	Range: 5 to 25
Max occupancy – Corridor-X47	Range: 100 to 200
Ventilation – Corridor-X48	Range: 5 to 25
Max occupancy – Laundry-X49	Range: 100 to 200
Ventilation – laundry-X50	Range: 15 to 35
Max occupancy – All others-X51	Range: 100 to 200
Ventilation -All others-X52	Range: 5 to 25

CHAPTER 4

DESIGN OF EXPERIMENTS

4.1 Kung's Design

The design of the experiments (DOE) is performed based on a mixed array (MA) [15] to handle discrete variables in the first experimental design. The adopted MA contains 96 runs, and 40, 1, and 16 variables with 2, 3, and 4 levels, respectively [36]. This MA design is selected to provide the closest MA design to the number of existing decision variables shown in Table 3.1, by limiting the number of levels of some of the variables based on the selected MA.

However, in order to adjust the levels of the variables, it is required to convert all of the variables with more than four levels into variables with only four levels. For example, "concrete slab on grade-X5" has six potential options of 2500, 3000, 4000, 5000, 6000, and 8000 psi in both software tools, i.e., eQUEST and ATHENA. Since it has more than four levels, only four of these levels will be considered. These levels were selected based on the equal spacing between the values, since there was no other preference in the level selection. Thus, only 3000, 4000, 5000, 6000 psi were considered and studied as the levels of concrete slabs in eQUEST and ATHENA. In addition, only one 3-level variable should exist in the model, therefore, only one of the 3-level variables is kept and the rest of them are converted into 2-level variables. The number of decision variables investigated in this study after performing the above mentioned adjustments is shown in Table 3.1. A 96-by-57 matrix is generated in the DoE.base package in R studio [37]. In this study, 23 two-level variables, one 3-level variable, and fourteen 4-level variables are investigated. Thus, only 23 columns of the 40 columns of this matrix that contain the 2-level variables are randomly assigned to the 2-level variables. In addition, 14 columns of the 16 columns for 4-level factors are randomly assigned to the 4-level variables.

The method explained in section 2.3.1 was used to generate the 192-point Kung's design, as shown in Table 4.1. The columns that are named MA and S represent the run number that has been selected from the mixed array and Sobol' sequence, respectively.

Table 4-1 Kung's Design for 192 points (runs) [14]

Runs #	MA	S	Runs #	MA	S	Runs #	MA	S	Runs #	MA	S	Runs #	MA	S
1	65	37	41	77	86	81	96	95	121	44	7	161	33	86
2	7	57	42	21	56	82	72	11	122	67	75	162	41	68
3	87	73	43	86	64	83	42	63	123	17	60	163	87	49
4	69	71	44	54	90	84	68	58	124	40	56	164	2	82
5	29	1	45	16	93	85	19	40	125	10	85	165	56	90
6	91	62	46	11	70	86	70	45	126	92	6	166	90	93
7	57	38	47	84	41	87	47	6	127	95	40	167	46	45
8	61	21	48	44	49	88	74	76	128	96	70	168	60	43
9	51	44	49	35	43	89	43	52	129	51	87	169	22	84
10	40	33	50	76	36	90	28	4	130	12	85	170	66	77
11	5	10	51	48	7	91	22	74	131	8	83	171	61	32
12	10	69	52	92	77	92	25	35	132	6	89	172	7	63
13	71	5	53	41	16	93	8	27	133	71	22	173	93	62
14	75	89	54	2	61	94	9	17	134	36	17	174	68	95
15	6	30	55	82	65	95	49	81	135	75	20	175	91	33
16	64	84	56	95	42	96	83	15	136	64	50	176	24	36
17	78	34	57	24	66	97	49	13	137	43	3	177	85	91
18	50	92	58	67	48	98	55	35	138	52	52	178	86	25
19	32	39	59	15	85	99	94	54	139	79	48	179	77	18
20	81	12	60	79	59	100	27	11	140	73	53	180	1	24
21	38	60	61	31	8	101	69	72	141	29	12	181	23	74
22	27	55	62	62	53	102	48	51	142	42	67	182	45	94
23	23	18	63	39	14	103	47	39	143	25	21	183	72	31
24	17	78	64	80	32	104	35	76	144	30	15	184	21	37
25	26	79	65	33	91	105	39	23	145	57	14	185	50	57

26	45	88	66	1	94	106	54	16	146	70	71	186	32	42
27	63	47	67	36	51	107	59	46	147	13	47	187	11	69
28	66	54	68	46	3	108	58	61	148	78	38	188	37	80
29	13	46	69	12	67	109	83	34	149	20	4	189	74	88
30	88	96	70	4	20	110	16	81	150	19	26	190	89	5
31	59	9	71	56	28	111	38	44	151	63	19	191	31	58
32	52	26	72	85	87	112	4	10	152	15	28	192	88	41
33	30	82	73	20	31	113	26	96	153	18	55			
34	60	50	74	94	72	114	14	1	154	62	27			
35	55	13	75	18	75	115	76	92	155	5	29			
36	3	68	76	14	22	116	3	2	156	80	8			
37	90	25	77	37	83	117	81	9	157	65	64			
38	93	19	78	34	80	118	82	73	158	53	66			
39	53	29	79	73	2	119	9	30	159	84	78			
40	58	23	80	89	24	120	28	59	160	34	79			

4.2 Martinez's Design

The second design is generated based on the method created by Martinez [18]. In this method, all of the variables, either categorical or numerical, are considered to be continuous. Discrete variables should be first scaled to the continuous space between zero and one using a Sobol' sequence with 192 runs. 52 decision variables are identified in this study to be investigated (see Table 3.1 and 3.4), with 38 discrete variables and 14 continuous variables. Among these 38 discrete variables, there are 23 variables with two levels, only one with three levels, and 14 variables with four levels. Since scaling all of the levels of a variable is redundant, for a discrete variable with p levels, only $p-1$ levels are scaled to continuous space, and it is not required to scale the last level. This means that a Sobol' sequence with 192 rows and 81 columns is required for this design. These 81 columns are randomly assigned to either the continuous variables or the levels of discrete variables.

As it was mentioned in literature review (Section 2.3.2), at first, it is needed to scale the numbers generated by Sobol' sequence to the respective range of each continuous variable, and then, use the generated numbers as the value of that specific continuous variable. Then, use the generated numbers as the value of that specific continuous variable. For discrete variables, as mentioned in section 2.3.2, it is required to calculate a threshold for either of factors based on the number of their levels to undo the scaling and convert them back into the discrete type.

4.3 Validation data set

Kung's method was also used to generate a new experimental design as the validation dataset for further analysis. This validation dataset includes a 96-by-14 matrix to handle the continuous variables based on Sobol' sequence approach, and a new 96-by-57 matrix to handle the discrete variables based on MA approach. These matrices are combined using the Latin hypercube method to make a validation dataset with 96-points [Appendix A]. Also, Martinez's design was used to generate a new experimental design with 96 runs as the validation dataset. This dataset includes a 96-by-81 matrix from Sobol's sequence to handle all of the predictors.

4.4 Response variables selection

As mentioned in Section 3, several responses or performance metrics are available from eQUEST and ATHENA. Since highly correlated response variables do not provide new information, an investigation is first performed to find the responses that are not highly correlated. In fact, the response variables that are highly correlated have the same pattern in estimation. By fitting a regression model to one of them, it is possible to have a prediction of the model of other correlated variables. Thus, to save time and effort, it is better to continue the analysis with those response variables that are not highly correlated. The scatter plot of response variables verses each other is shown in Figure 4.1, and the correlation coefficient between them is shown in Table 4.2. These results are the outputs of the 192 simulations in the two software tools.

To make it easier to analyze the response variables, the twelve response variables can be categorized into three separate groups. The first group includes the first three variables which are from eQUEST, and, as it can be seen, a clear linear trend is seen in the scatter plot, which shows they are highly correlated. The correlation coefficient between these three variables is 0.91295. The rest of the nine variables are the results of the simulations in ATHENA. These nine variables also can be divided into two highly correlated groups; thus, the second group includes the first six performance metrics, from GWP to Smog-pot, which show a clear linear relationship in the scatter plot and a high correlation coefficient. In fact, the smallest correlation coefficient between the first group of variables is 0.83212. On the other hand, the last three performance metrics from ATHENA that can be categorized as the third group of response variables demonstrate an obvious linear trend that can be recognized in their scatter plot. The smallest correlation coefficient between them is 0.99493. It is important to note that the responses from eQUEST and ATHENA have a very low correlation with each other, which is expected. In addition, the biggest correlation coefficient between the two groups of response variables from ATHENA is 0.77355. Thus, among the three response variables from eQUEST (from the first group), only “annual source energy-total” in million British thermal unit (Mbtu), and among the nine response variables from ATHENA, only two response variables “Global Warming Potential (GWP)” in kg CO_2 equivalent mass from second group and “non-renewable energy” in mega joule (MJ) from third group have been selected for further analysis.

Table 4-2 Pearson Correlation Coefficient between twelve response variables from eQUEST and ATHENA for the first design

Pearson Correlation Coefficients, N = 192
Prob > |r| under H0: Rho=0

	Anl_src_energy	Anl_energy_usage	HVAC	GWP	Acid_pot	HH_pot	EUt_pot	Ozone_Pot	Smog_pot	Primary_energy	Non_renewable_energy	Fossil_fuel
Anl_src_energy	1.00000	0.91295 <.0001	0.91295 <.0001	0.00651 0.9286	0.03208 0.6587	0.02862 0.6935	-0.00370 0.9594	-0.01286 0.8594	-0.02444 0.7365	-0.02332 0.7482	-0.02179 0.7642	-0.03070 0.6725
Anl_energy_usage	0.91295 <.0001	1.00000	1.00000 <.0001	0.04760 0.5121	0.07872 0.2778	0.07434 0.3055	0.02839 0.6959	0.02713 0.7088	0.01966 0.7866	-0.00808 0.9114	-0.00870 0.9047	-0.02014 0.7816
HVAC	0.91295 <.0001	1.00000 <.0001	1.00000	0.04760 0.5121	0.07871 0.2778	0.07434 0.3054	0.02839 0.6958	0.02713 0.7088	0.01966 0.7866	-0.00810 0.9112	-0.00872 0.9045	-0.02016 0.7814
GWP	0.00651 0.9286	0.04760 0.5121	0.04760 0.5121	1.00000	0.96787 <.0001	0.91679 <.0001	0.97428 <.0001	0.95272 <.0001	0.97319 <.0001	0.77355 <.0001	0.76732 <.0001	0.71717 <.0001
Acid_pot	0.03208 0.6587	0.07872 0.2778	0.07871 0.2778	0.96787 <.0001	1.00000	0.89079 <.0001	0.90845 <.0001	0.89240 <.0001	0.92555 <.0001	0.75472 <.0001	0.74635 <.0001	0.69678 <.0001
HH_pot	0.02862 0.6935	0.07434 0.3055	0.07434 0.3054	0.91679 <.0001	0.89079 <.0001	1.00000	0.87049 <.0001	0.83212 <.0001	0.89378 <.0001	0.62304 <.0001	0.61406 <.0001	0.55909 <.0001
EUt_pot	-0.00370 0.9594	0.02839 0.6959	0.02839 0.6958	0.97428 <.0001	0.90845 <.0001	0.87049 <.0001	1.00000	0.96817 <.0001	0.97695 <.0001	0.74284 <.0001	0.73457 <.0001	0.68585 <.0001
Ozone_Pot	-0.01286 0.8594	0.02713 0.7088	0.02713 0.7088	0.95272 <.0001	0.89240 <.0001	0.83212 <.0001	0.96817 <.0001	1.00000	0.94022 <.0001	0.75641 <.0001	0.75248 <.0001	0.70814 <.0001
Smog_pot	-0.02444 0.7365	0.01966 0.7866	0.01966 0.7866	0.97319 <.0001	0.92555 <.0001	0.89378 <.0001	0.97695 <.0001	0.94022 <.0001	1.00000	0.72634 <.0001	0.71745 <.0001	0.67147 <.0001
Primary_energy	-0.02332 0.7482	-0.00808 0.9114	-0.00810 0.9112	0.77355 <.0001	0.75472 <.0001	0.62304 <.0001	0.74284 <.0001	0.75641 <.0001	0.72634 <.0001	1.00000	0.99894 <.0001	0.99493 <.0001
Non_renewable_energy	-0.02179 0.7642	-0.00870 0.9047	-0.00872 0.9045	0.76732 <.0001	0.74635 <.0001	0.61406 <.0001	0.73457 <.0001	0.75248 <.0001	0.71745 <.0001	0.99894 <.0001	1.00000	0.99668 <.0001
Fossil_fuel	-0.03070 0.6725	-0.02014 0.7816	-0.02016 0.7814	0.71717 <.0001	0.69678 <.0001	0.55909 <.0001	0.68585 <.0001	0.70814 <.0001	0.67147 <.0001	0.99493 <.0001	0.99668 <.0001	1.00000

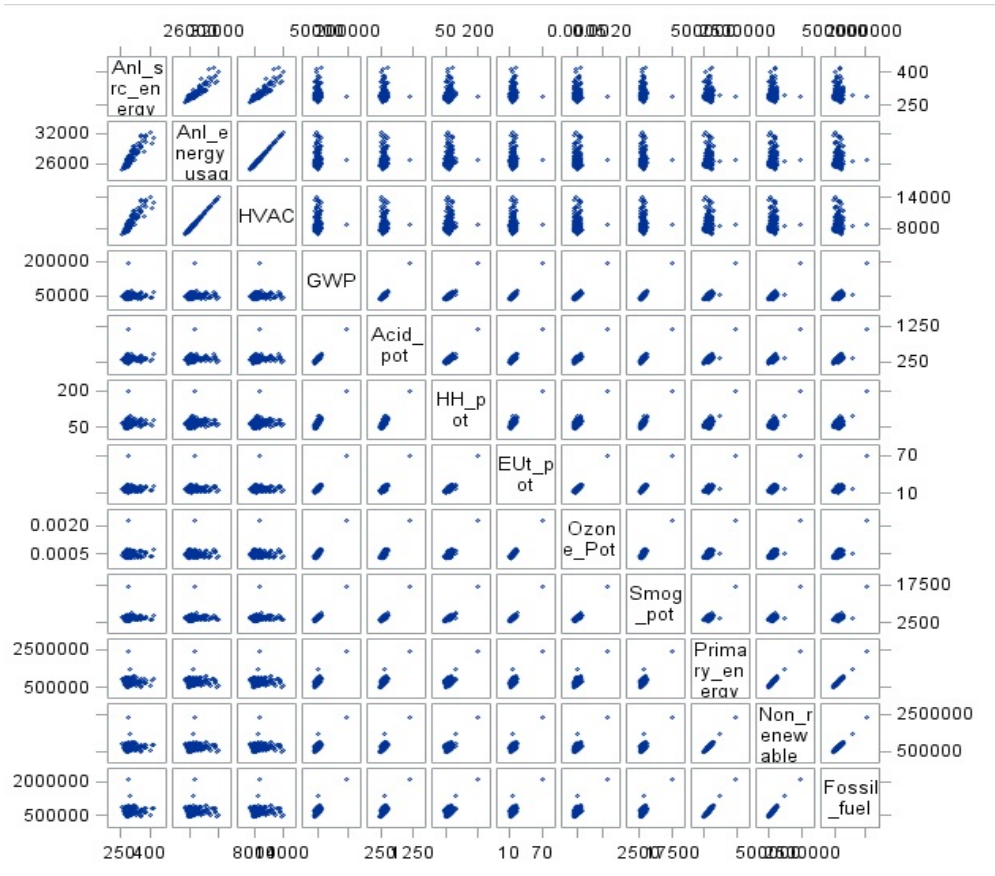


Figure 4.1 Scatter plot of twelve response variables from eQUEST and ATHENA for the first design

For Martinez’s design, similar to Kung’s design, three main groups are recognized among these twelve performance metrics based on correlation between them as follows; the first group includes all of the three response variables from eQUEST that are highly correlated, however, the correlation is not as strong as the correlation in Kung’s design. This is because the correlation coefficient between these three performance metrics is in the range of 0.69483 and 0.90571. In fact, the first and the third eQUEST outputs have the highest correlation with each other among all three performance metrics, and they have a small correlation with the second one (>0.6948). Meanwhile, these three outputs have a small correlation with the outputs from ATHENA, which is expected (<0.16466).

The second group includes the first six performance metrics from ATHENA. The highest correlation coefficient between these performance metrics is 0.93022, while the lowest correlation is 0.38517 and is related to “eutrophication potential.”.” However, the correlation between “eutrophication potential” and other outputs in this group is not high; this performance metric has the highest correlation coefficient with the metrics in this group rather than other outputs. In other words, as it can be seen in Table 5.1, the correlation coefficient between “eutrophication potential” and other outputs (the outputs that are not considered in the second group of correlated performance metrics) is small (<0.2078). Thus, “eutrophication potential” can be kept in the second group of the highly correlated performance metrics. The third group includes the last three outputs from ATHENA, “non-renewable energy,” “primary energy,” and “Fossil Fuel Consumption.” These three responses are highly correlated (correlation coefficient>0.99), meanwhile, there is a small correlation between these three performance metrics and other ones.

Accordingly, three different groups are identified from the highly correlated response variables. The first group includes “annual source energy,” “annual site energy,” and “HVAC.” “GWP,” “acidification potential,” and “HH potential,” “eutrophication potential,” “ozone depletion potential,” and “photochemical smog potential” are in the second group. “Non-renewable energy,” “primary energy,” and “Fossil Fuel Consumption” are categorized in the third group. The scatter plots presented in Figure 5.1 confirms the above results.

For Kung’s design, three performance metrics, i.e., “annual source energy,” “GWP,” and “non-renewable energy,” were selected to be investigated. Since the main goal of this study was to compare the two designs, it is preferred to keep the same performance metrics in both. Thus, it was decided to continue with “annual source energy,” “GWP,” and “non-renewable energy” for further analysis for Martinez’s design. Moreover, the result of the recent analysis confirms that the selection of these three performance metrics, which are not highly correlated based on Martinez’s design, is reasonable and they can be used for future analyses.

Table 4-3 Pearson Correlation Coefficient between twelve response variables from eQUEST and ATHENA for Martinez's design

Pearson Correlation Coefficients, N = 192
Prob > |r| under H0: Rho=0

	Anl_Energy_Usage	Anl_Site_energy	HVAC	GWP	Acid_pot	HH_pot	EUt_pot	Ozone_Pot	Smog_pot	Primary_energy	Non_renewable_energy	Fossil_fuel
Anl_Energy_Usage	1.00000	0.69483 <.0001	0.90571 <.0001	0.03980 0.5836	0.14036 0.0522	-0.00063 0.9930	-0.04201 0.5629	0.05574 0.4425	0.00130 0.9857	0.04893 0.5003	0.04532 0.5325	0.04531 0.5326
Anl_Site_energy	0.69483 <.0001	1.00000	0.72303 <.0001	0.07759 0.2848	0.15391 0.0331	-0.00800 0.9123	0.02505 0.7302	0.13392 0.0640	0.05574 0.4426	0.10852 0.1341	0.10554 0.1451	0.10724 0.1387
HVAC	0.90571 <.0001	0.72303 <.0001	1.00000	0.05389 0.4579	0.16466 0.0225	0.02777 0.7022	-0.03366 0.6430	0.05384 0.4583	0.00372 0.9592	0.09134 0.2077	0.08590 0.2361	0.08719 0.2292
GWP	0.03980 0.5836	0.07759 0.2848	0.05389 0.4579	1.00000	0.91486 <.0001	0.83994 <.0001	0.42935 <.0001	0.84556 <.0001	0.93022 <.0001	0.54890 <.0001	0.53531 <.0001	0.46046 <.0001
Acid_pot	0.14036 0.0522	0.15391 0.0331	0.16466 0.0225	0.91486 <.0001	1.00000	0.76566 <.0001	0.38517 <.0001	0.71243 <.0001	0.81199 <.0001	0.54056 <.0001	0.52441 <.0001	0.45679 <.0001
HH_pot	-0.00063 0.9930	-0.00800 0.9123	0.02777 0.7022	0.83994 <.0001	0.76566 <.0001	1.00000	0.30619 <.0001	0.60411 <.0001	0.79605 <.0001	0.20786 0.0038	0.19211 0.0076	0.11156 0.1234
EUt_pot	-0.04201 0.5629	0.02505 0.7302	-0.03366 0.6430	0.42935 <.0001	0.38517 <.0001	0.30619 <.0001	1.00000	0.59199 <.0001	0.43957 <.0001	0.20643 0.0041	0.19200 0.0076	0.16526 0.0220
Ozone_Pot	0.05574 0.4425	0.13392 0.0640	0.05384 0.4583	0.84556 <.0001	0.71243 <.0001	0.60411 <.0001	0.59199 <.0001	1.00000	0.81402 <.0001	0.53435 <.0001	0.52304 <.0001	0.46795 <.0001
Smog_pot	0.00130 0.9857	0.05574 0.4426	0.00372 0.9592	0.93022 <.0001	0.81199 <.0001	0.79605 <.0001	0.43957 <.0001	0.81402 <.0001	1.00000	0.41557 <.0001	0.39666 <.0001	0.33003 <.0001
Primary_energy	0.04893 0.5003	0.10852 0.1341	0.09134 0.2077	0.54890 <.0001	0.54056 <.0001	0.20786 0.0038	0.20643 0.0041	0.53435 <.0001	0.41557 <.0001	1.00000	0.99839 <.0001	0.99287 <.0001
Non_renewable_energy	0.04532 0.5325	0.10554 0.1451	0.08590 0.2361	0.53531 <.0001	0.52441 <.0001	0.19211 0.0076	0.19200 0.0076	0.52304 <.0001	0.39666 <.0001	0.99839 <.0001	1.00000	0.99568 <.0001
Fossil_fuel	0.04531 0.5326	0.10724 0.1387	0.08719 0.2292	0.46046 <.0001	0.45679 <.0001	0.11156 0.1234	0.16526 0.0220	0.46795 <.0001	0.33003 <.0001	0.99287 <.0001	0.99568 <.0001	1.00000

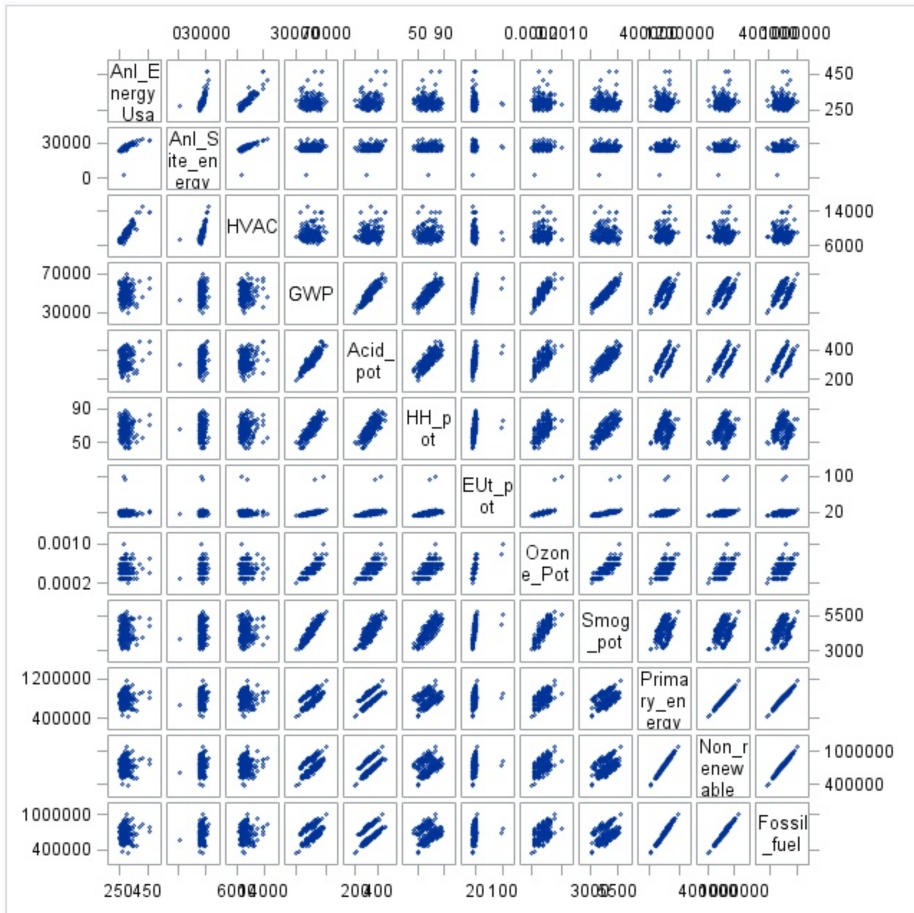


Figure 4.2 Scatter plot of twelve response variables from eQUEST and ATHENA for Martinez's design

CHAPTER 5

STATISTICAL MODELS FOR GREEN BUILDING PERFORMANCE OUTPUTS

Treed regression model is one of the statistical analysis methods used in this study. In order to apply the treed regression model, first, it is required to determine the regression trees. Thus, in the first section of this chapter, the settings used for the tree generation are described. Fitting the tree models are described in the second section, which is followed by an explanation of the fitting of the regression models at the terminal nodes of each tree.

5.1 Determining the setting for tree

5.1.1 Kung's Design

In order to determine the regression trees in the CART module from Salford system [24], a sample size of 35 was selected in the "controls for tree growth due to sample size" setting. By limiting the sample size, the software will prevent the generation of TNs with fewer than 35 cases in each TN. The trees can be generated using all of the predictors or based only on the categorical variables. Therefore, it is important to investigate if there is any difference between the outcomes of the two scenarios. In addition, it is required to select some thresholds for splitting the nodes in each tree. In this study, two limits are selected to use in the tree generation in each method. The remaining factors are set as default in CART [24].

The minimum number of cases in terminal nodes (TN) is plotted versus the limit for the number of cases in each node to be split in Figure 5.1-5.5 and is presented in Table 5.1-5.5. In Table 5.1-5.5, the first column is showing the limit for number of the cases in each TN to be split. This is the number that can be controlled in this study. The rest of the columns are the output from CART based on the limit given to the CART. The second column indicates the number of the TNs. The third column is the minimum number of cases in TN. The fourth column shows the relative error and the last column indicates the difference between the current limit and the previous one. The selection of the limits to be investigated can be based on three different factors, 1) biggest difference between the relative errors, 2) the biggest difference between minimum number of the cases in TNs, and 3) number of TN. This can be further explored from Table 5.1-5.5.

Table 5-1 Trees summary based on “annual source energy” and when considering all predictors for Kung’s design

limit	Node	Min node cases	Relative error	Diff. error
40	6	20	0.411	---
45	7	22	0.413	-0.002
50	5	31	0.454	-0.041
55	5	31	0.454	0
60	5	31	0.454	0
65	5	33	0.489	-0.035
67	5	33	0.489	0
68	5	35	0.462	0.027
70	5	35	0.462	0
76	5	35	0.462	0
78	4	35	0.483	-0.021
79	3	37	0.492	-0.009
80	3	37	0.492	0
85	3	37	0.492	0
90	3	37	0.492	0

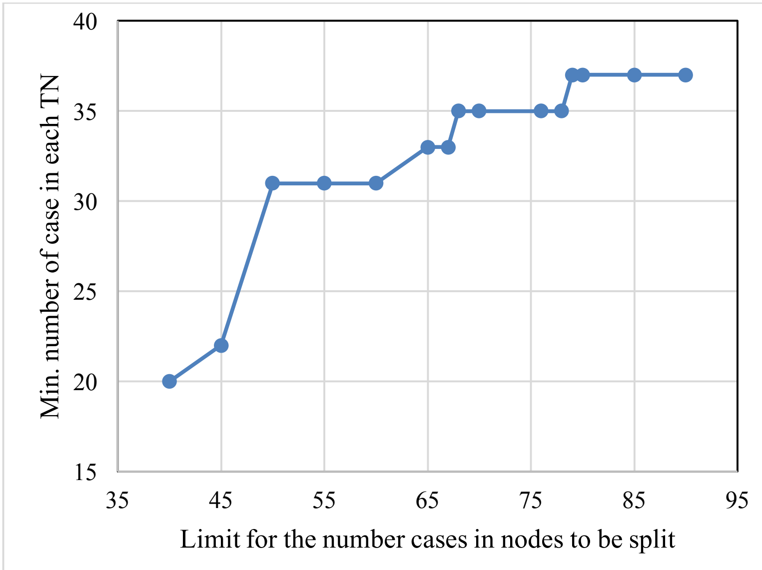


Figure 5.1 The minimum number of cases in each TN vs. the limit for the number of cases in nodes to be split in the tree based on “annual source energy” and considering all variables for Kung’s design

Table 5-2 Trees summary based on “GWP” and when considering all predictors for Kung’s design

Limit	Node	Min node cases	Relative error	Diff. error
40	2	96	0.51	---
45	2	96	0.51	0
50	2	96	0.51	0
51	2	96	0.51	0
52	3	48	0.468	0.042
55	3	48	0.468	0
60	3	48	0.468	0
65	3	48	0.468	0
70	3	48	0.468	0
75	3	48	0.468	0
80	3	48	0.468	0
85	3	48	0.468	0
90	3	48	0.468	0

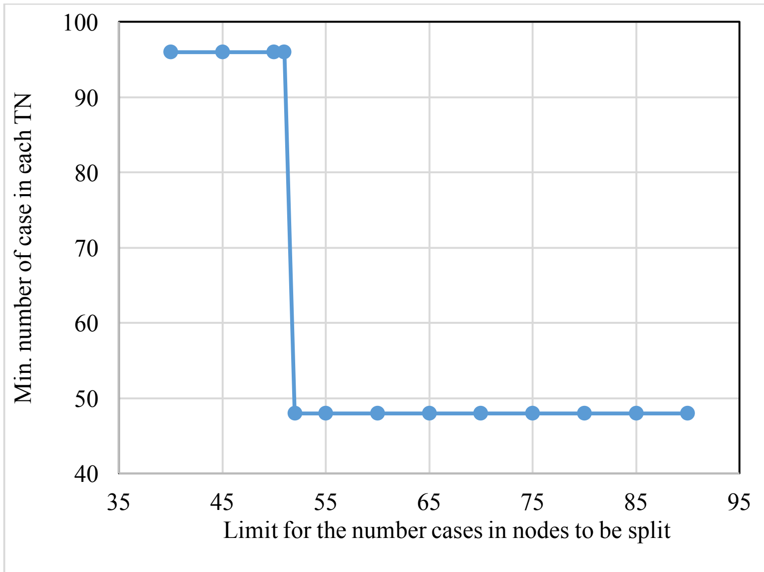


Figure 5.2 The minimum number of cases in each TN vs. the limit for the number of cases in nodes to be split in the tree based on “GWP” and considering all variables for Kung’s design

Table 5-3 Trees summary based on “GWP” and when considering only the categorical predictors for Kung’s design

limit	Node	Min node cases	Relative error	Diff. error
40	2	64	0.983	---
45	2	64	0.983	0
50	2	64	0.983	0
55	2	64	0.983	0
60	2	64	0.983	0
65	2	64	0.983	0
70	2	64	0.983	0
75	2	64	0.983	0
80	2	64	0.983	0
85	2	64	0.983	0
90	2	64	0.983	0

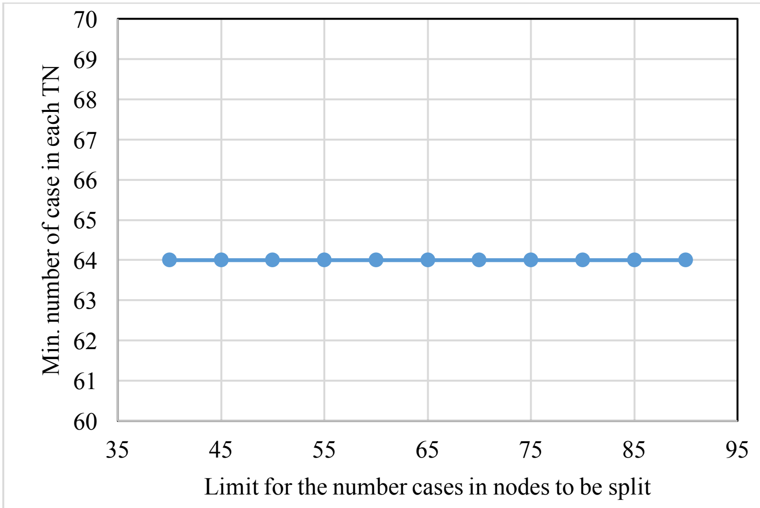


Figure 5.3 The minimum number of cases in each TN vs. the limit for the number of cases in nodes to be split in the tree based on "GWP" and considering only categorical variables for Kung's design

Table 5-4 Trees summary based on "non-renewable energy" and when considering all the predictors for Kung's design

limit	Node	Min node cases	Relative error	Diff. error
40	5	32	0.52	---
45	5	32	0.52	0
50	5	32	0.52	0
55	5	32	0.52	0
60	5	32	0.52	0
64	5	32	0.52	0
65	4	32	0.582	-0.062
70	3	64	0.589	-0.007
75	3	64	0.589	0
80	3	64	0.589	0
85	3	64	0.589	0
90	3	64	0.589	0

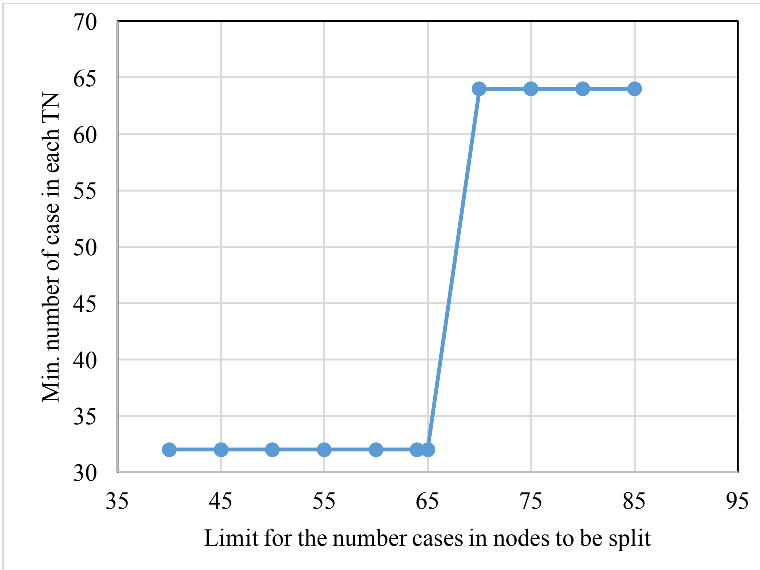


Figure 5.4 The minimum number of cases in each TN vs. the limit for the number of cases in nodes to be split in the tree based on “non-renewable energy” considering all variables for Kung’s design

Table 5-5 Trees summary based on “non-renewable energy” and when considering only the categorical predictors for Kung’s design

limit	Node	Min node cases	Relative error	Diff. error
40	4	32	0.568	---
45	4	32	0.568	0
50	4	32	0.568	0
55	4	32	0.568	0
60	4	32	0.568	0
64	4	32	0.568	0
65	3	64	0.589	-0.021
66	3	64	0.589	0
70	3	64	0.589	0
75	3	64	0.589	0
80	3	64	0.589	0
85	3	64	0.589	0
90	3	64	0.589	0

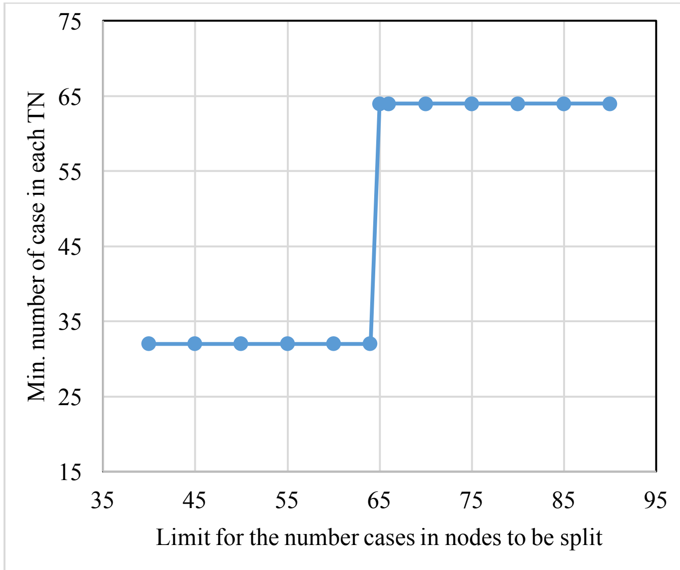


Figure 5.5 The minimum number of cases in each TN vs. the limit for the number of cases in nodes to be split in the tree based on “non-renewable energy” and considering only categorical variables for Kung’s design

The trees summary for the “annual source energy” and considering all of the variables, are shown in Figure 5.1 and Table 5.1. Since the next step after fitting the trees is to fit the regression models at terminal nodes (TN), the number of cases in each TN should be traced closely. It is preferred not to have less than 30 cases as the number of cases in each TN. Thus, the limits of 40 and 45 for splitting are ignored here, however based on Table 5.1 the biggest jump is between these two limits. In addition, as it can be seen in the plot, the first biggest difference between relative errors is between the limits of 60 and 65, and the second biggest difference is between 67 and 68. Since by changing the limits from 60 into 65 the relative error becomes bigger (which is not desired), the limits 67 and 68 has been selected as two cases to be investigated when the tree is generated based on “annual source energy” and all of the predictors. Salford Systems’ CART software does not add any predictor variables to generate a tree for “annual source energy,” when the tree is generated by considering categorical variables only. Thus, this case will not be considered in the CART method.

According to Figure 5.2 and Table 5.2, for “GWP” and considering all of the variables, a downward trend can be seen in the plot of the minimum number of cases in each TN vs. the limit for the number of cases in nodes to be split in the tree based on “GWP” and considering all of the variables; this downward trend is unexpected. In fact, it is expected to see an upward trend in this plot, since the minimum number of cases in each TN is expected to increase by increasing the limit for the number of cases in nodes to be split in the tree. Although there is such an unexpected downward trend in the plot, a big jump is seen between the minimum numbers of cases in each TN when the limit for splitting nodes changes from 51 into 52 cases. Thus, these two limits have been selected to be investigated for this study when the tree is generated based on “GWP” and all of the predictors. In addition, it seems that by changing the limit for the number of cases in nodes to be split in the tree based on “GWP” and considering only categorical variables, Figure 5.3 and Table 5.3, the minimum number of cases in each TN does not change. Thus, there is no priority in selecting the limits to split the node here to investigate, and one of these limits, i.e., 60 cases, has been selected to be investigated.

According to the plot of the minimum number of cases in each TN vs. the limit for the number of cases in nodes to be split in the tree based on “non-renewable energy” and considering all of the variables (Figure 5.4 and Table 5.4), there is a big difference between the minimum number of cases in each TN when the limit for split the nodes changes from 64 into 65 cases. Thus, for this study these two situations have been selected to be investigated.

Finally, according to the plot of the minimum number of cases in each TN vs. the limit for the number of cases in nodes to be split in the tree based on “non-renewable energy” and considering only the categorical variables (Figure 5.5 and Table 5.5), there is a big difference between the minimum number of cases in each TN when the limit for splitting the nodes changes from 64 into 65 cases. Thus, for this study these two situations have been selected to be investigated. Thus, based on Kung’s design, the limits to split the nodes for each response to be investigated in this study are as follows:

For “annual source energy” and considering all of the variables the limits are 67 and 68 cases. These limits for “GWP” and considering all of the variables, are 51 and 52 cases. The limit for “GWP” and only the categorical variables considered is only the limit 60 cases. In addition, these limits for “non-renewable energy” and regardless of the type of the variables are 64 and 65 cases.

5.1.2 Martinez’s Design

In this step, Salford Systems software was used to build trees model for Martinez’s design. Since it is required in this study to have similar assumptions, the same assumptions used in Kung’s design in Section 4.3 were considered in CART module. Table 5.6-5.11 and Figure 5.6-5.11 are shown the trees summary for all of the responses.

Table 5-6 Trees summary based on “annual source energy” and when considering all predictors for Martinez’s design

Limit	Node	Minimum node cases	Relative error	Diff. error
40	3	28	0.451	---
45	3	28	0.451	0
50	3	28	0.451	0
55	3	28	0.451	0
57	3	28	0.451	0
58	3	35	0.416	0.035
60	3	35	0.416	0
65	3	35	0.416	0
70	3	35	0.416	0
75	3	35	0.416	0
80	3	35	0.416	0
85	3	35	0.416	0
90	3	35	0.416	0

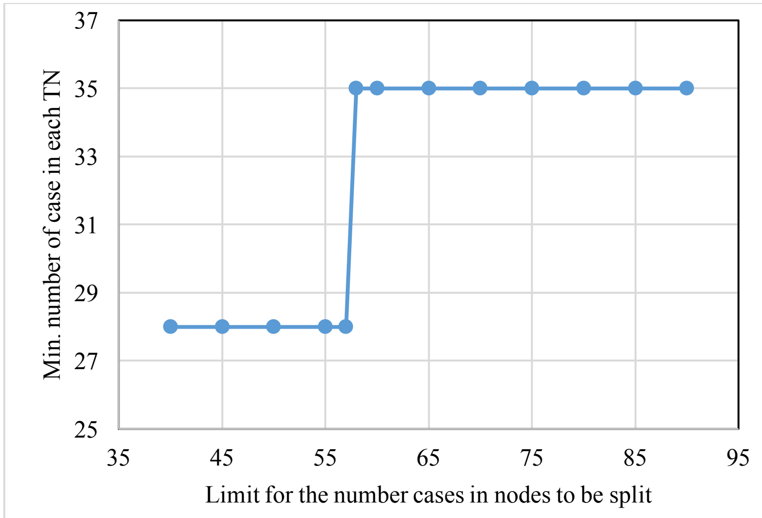


Figure 5.6 The minimum number of cases in each TN vs. the limit for the number of cases in nodes to be split in the tree based on “annual source energy” and considering all variables for Martinez’s design

Table 5-7 Trees summary based on “annual source energy” and when considering only categorical predictors for Martinez’s design

Limit	Node	Minimum node cases	Relative error	Diff. error
40	2	91	0.998	---
45	2	91	0.998	0
50	2	91	0.998	0
55	2	91	0.998	0
60	2	91	0.998	0
65	2	91	0.998	0
70	2	91	0.998	0
75	2	91	0.998	0
80	2	91	0.998	0
85	2	91	0.998	0
90	2	91	0.998	0

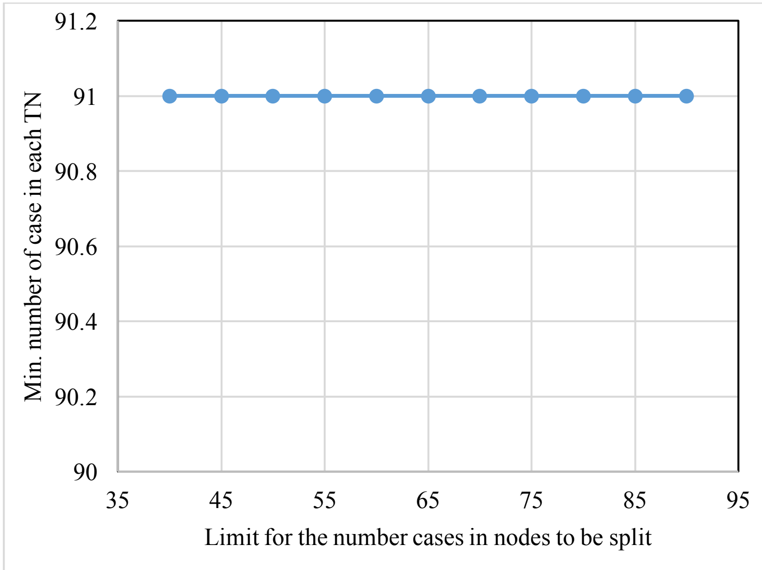


Figure 5.7 The minimum number of cases in each TN vs. the limit for the number of cases in nodes to be split in the tree based on “annual source energy” and considering only the categorical variables for Martinez’s design

Table 5-8 Trees summary based on “GWP” and when considering all predictors for Martinez’s design

Limit	Node	Minimum node cases	Relative error	Diff. error
40	7	22	0.329	---
45	7	22	0.329	0
50	5	30	0.356	-0.027
55	5	30	0.356	0
60	5	30	0.356	0
65	4	34	0.388	-0.032
69	4	34	0.388	0
70	3	51	0.356	0.032
75	3	51	0.356	0
80	3	51	0.356	0
85	3	51	0.356	0
90	3	51	0.356	0

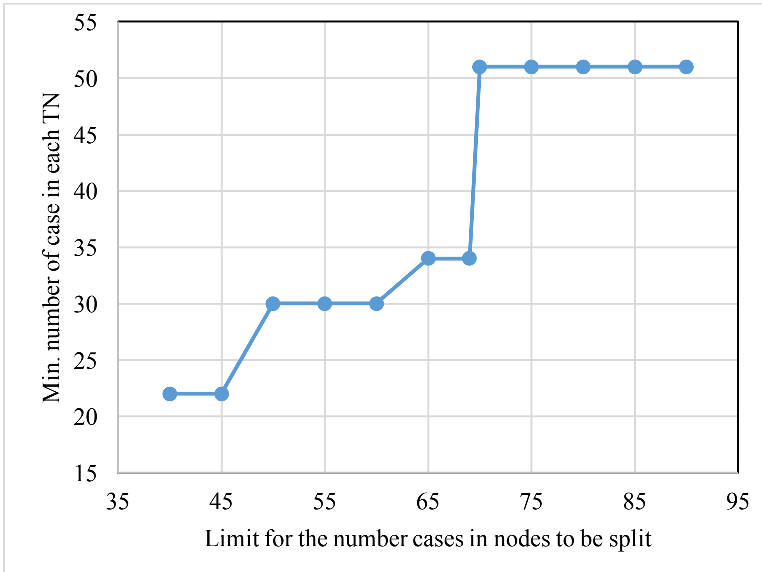


Figure 5.8 The minimum number of cases in each TN vs. the limit for the number of cases in nodes to be split in the tree based on “GWP” and considering all variables for Martinez’s design

Table 5-9 Trees summary based on “GWP” and when considering only categorical predictors for Martinez’s design

<table>
<tr><th>Limit</th><th>Node</th><th>Minimum node cases</th><th>Relative error</th><th>Diff. error</th></tr>
<tr><td>40</td><td>3</td><td>33</td><td>0.975</td><td>---</td></tr>
<tr><td>45</td><td>3</td><td>33</td><td>0.975</td><td>0</td></tr>
<tr><td>50</td><td>3</td><td>33</td><td>0.975</td><td>0</td></tr>
<tr><td>55</td><td>3</td><td>33</td><td>0.975</td><td>0</td></tr>
<tr><td>60</td><td>3</td><td>33</td><td>0.975</td><td>0</td></tr>
<tr><td>65</td><td>3</td><td>33</td><td>0.975</td><td>0</td></tr>
<tr><td>67</td><td>3</td><td>33</td><td>0.975</td><td>0</td></tr>
<tr><td>68</td><td colspan="4" rowspan="6">no tree</td></tr>
<tr><td>70</td></tr>
<tr><td>75</td></tr>
<tr><td>80</td></tr>
<tr><td>85</td></tr>
<tr><td>90</td></tr>
</table>

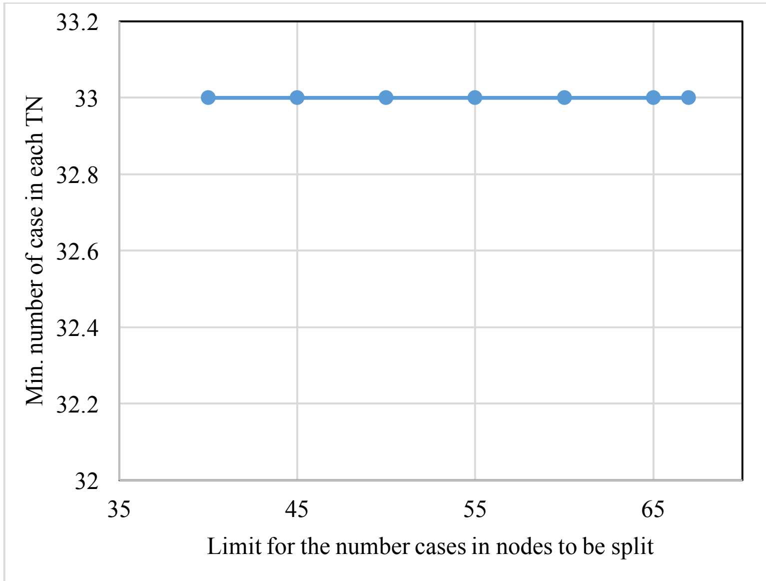


Figure 5.9 The minimum number of cases in each TN vs. the limit for the number of cases in nodes to be split in the tree based on “GWP” and considering only categorical predictors for Martinez’s design

Table 5-10 Trees summary based on “non-renewable energy” and when considering all predictors for Martinez’s design

Limit	Node	Minimum node cases	Relative error	Diff. error
40	6	24	0.253	---
45	6	24	0.253	0
50	5	34	0.314	-0.061
55	5	34	0.314	0
60	5	34	0.314	0
65	5	34	0.314	0
70	5	35	0.404	-0.09
72	5	35	0.404	0
73	3	48	0.458	-0.054
75	3	48	0.458	0
80	3	48	0.458	0
85	3	48	0.458	0
90	3	48	0.458	0

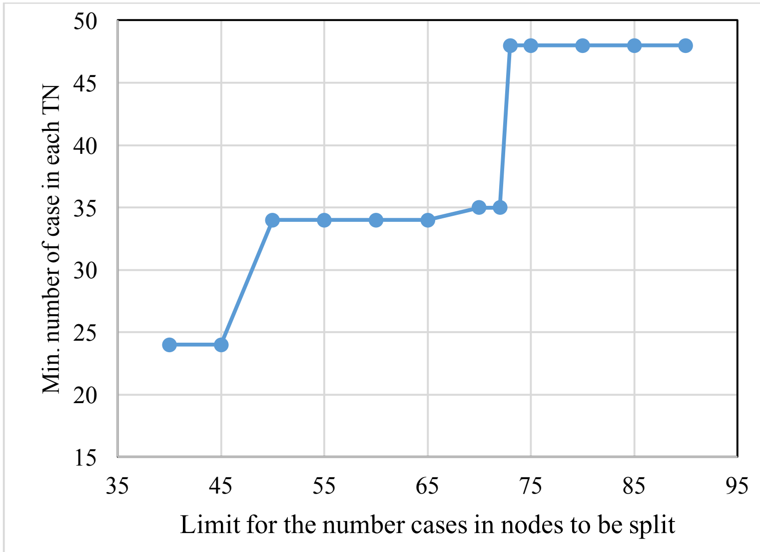


Figure 5.10 The minimum number of cases in each TN vs. the limit for the number of cases in nodes to be split in the tree based on "non-renewable energy" and considering all predictors for Martinez's design

Table 5-11 Trees summary based on "non-renewable energy" and when considering only categorical predictors for Martinez's design

Limit	Node	Minimum node cases	Relative error	Diff. error
40	3	59	0.413	---
45	3	59	0.413	0
50	3	59	0.413	0
55	3	59	0.413	0
60	3	59	0.413	0
65	3	59	0.413	0
66	4	35	0.412	0.001
70	4	35	0.412	0
72	4	35	0.412	0
73	3	59	0.413	-0.001
75	3	59	0.413	0
80	3	59	0.413	0
85	3	59	0.413	0
90	3	59	0.413	0

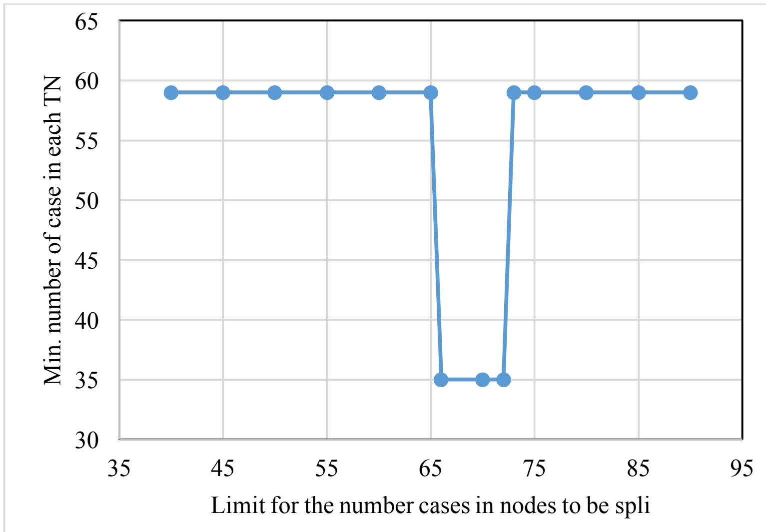


Figure 5.11 The minimum number of cases in each TN vs. the limit for the number of cases in nodes to be split in the tree based on “non-renewable energy” and considering only categorical predictors for Martinez’s design

Based on Figure 5.6 and Table 5.6, for “annual source energy” and when the tree is based on all of the predictors, the only difference between relative errors is between the limits of 57 and 58 cases. By changing the limit from 57 into 58, the minimum number of samples in each TN changes from 28 into 35, while the number of TNs does not change (=3). Thus, these two limits are selected to be investigated for this performance metric, when all of the predictors are considered in the tree generation based on Martinez’s design. Based on the results for “annual source energy” in Martinez’s design, the CART module can handle the predictor variables and generate the tree by considering categorical variables only, which is not in accordance with Kung’s design. In Table 5.7 and Figure 5.7, it seems that by changing the limit for the number of cases in nodes to be split in the tree based on “annual source energy” and considering only categorical variables, the minimum number of cases in each TN does not change. Thus, there is no priority in selecting the limits to split the node here to investigate, and one of these limits, i.e., 60 cases, has been selected to be investigated.

According to Figure 5.8 and Table 5.8, for “GWP” and considering all of the variables, the biggest jump between the minimum number of cases in each TN (=17) happens when the limit for splitting nodes changes from 69 into 70 cases. In addition, the relative error has the biggest change (=0.032) when the limit to split the nodes changes from 69 to 70 cases. Thus, these two limits have been selected to be investigated for this study, when the tree is generated based on “GWP” and all of the predictors. In addition, it seems that by changing the limit for the number of cases in nodes to be split in the tree based on “GWP” and considering only categorical variables from 67 to the bigger numbers (see Figure 5.9 and Table 5.9) The CART module does not add any predictor variables to generate a tree for “GWP.” In addition, by changing the limit between 67 cases and any number less than 67, for the number of cases in nodes to be split in the tree based on “GWP” and considering only categorical variables, the minimum number of cases in each TN does not change. Thus, there is no priority in selecting the limits to split the nodes, and one of these limits, i.e., 60 cases, has been selected to be investigated.

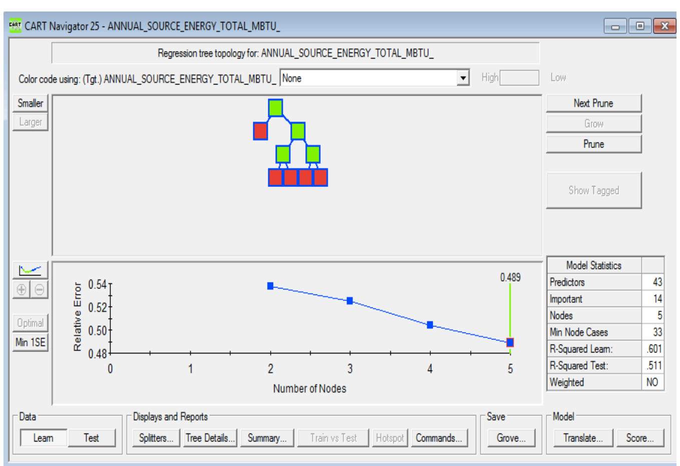
Based on Table 5.10 and Figure 5.10, for “non-renewable energy” and considering all of the variables, the biggest jump between the minimum number of cases in each TN (=13) happens when the limit for splitting nodes changes from 72 into 73 cases. Although the biggest difference between relative errors is given by changing the limit to split the nodes from 65 to 70, the minimum number of cases in each TN changes only one case. Thus, for this study, 72 and 73 cases are selected to be investigated for “non-renewable energy,” when the tree generation is based on all of the predictors. In addition, for “non-renewable energy,” when only categorical variables are considered to generate the tree, by changing the limit for the number of cases in nodes to be split in the tree, the same tree is given, except for limits from 66 to 72 cases (see Table 5.11 and Figure 5.11). The tree based on the limit 66 to 72 cases for the number of cases in nodes to be split has 4 TNs with 35 minimum number of samples. The relevant error is 0.413, which is not so different with the relative error for the trees with 3 TNs (=0.412). Thus, in this situation, 72 and 73 cases are selected to be investigated. Thus, based on Martinez's design, the limits to split the nodes for each response to be investigated in this study are as follows:

For "annual source energy" and considering all of the variables the limits are 57 and 58 cases. For "annual source energy" and only considering the categorical variables, only the limit 60 cases is investigated. These limits for "GWP" and considering all of the variables, are 69 and 70 cases. The limit for "GWP" and only the categorical variables considered is only the limit 60 cases. In addition, these limits for "non-renewable energy" and regardless of the type of the variables are 72 and 73 cases.

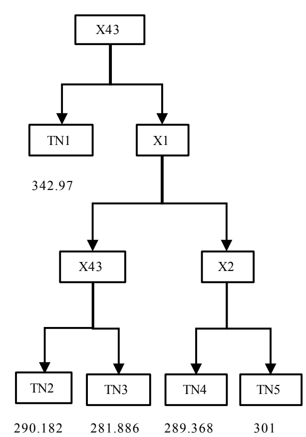
5.2 Fitting the Tree Models

5.2.1 Kung's Design

The trees generated in the CART module based on the selected thresholds for each of the above-mentioned cases are shown in Figure 5.12 to 5.20. Table 5.12 summarizes the tree generation results obtained from the CART module by considering all of the predictors to generate the trees or by considering categorical variables only.

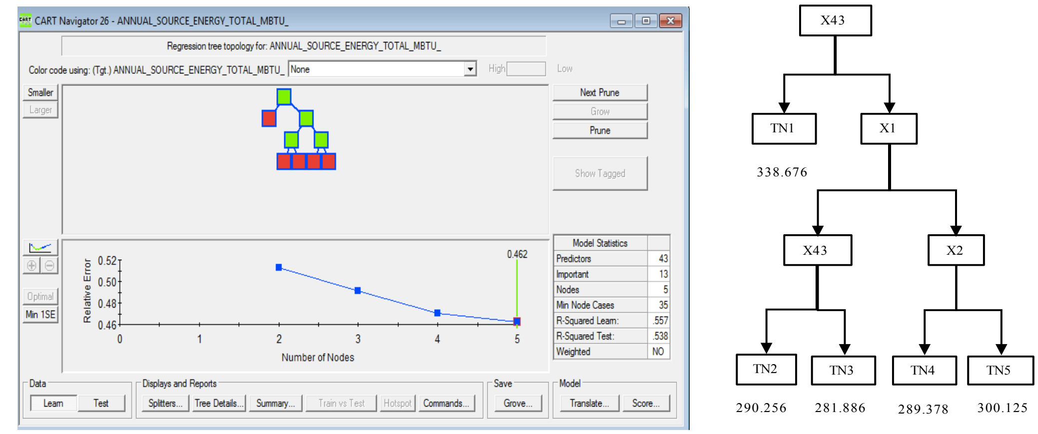


(a)



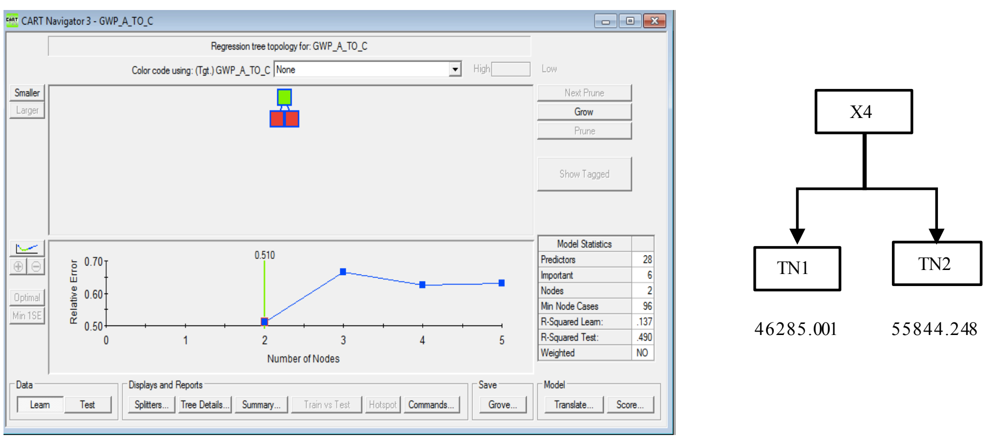
(b)

Figure 5.12 a) Tree details; b) Tree model for “annual source energy” when all of the predictor variables are considered and the limit of splitting the nodes is “not split if there is less than 67 cases” based on Kung’s design



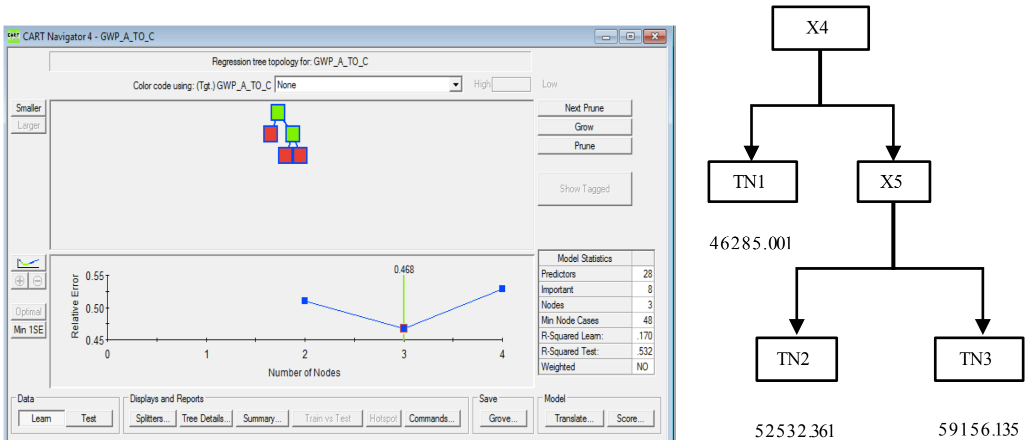
(a) (b)

Figure 5.13 a) Tree details; b) Tree model for “annual source energy” when all of the predictor variables are considered and the limit of splitting the nodes is “not split if there is less than 68 cases” based on Kung’s design



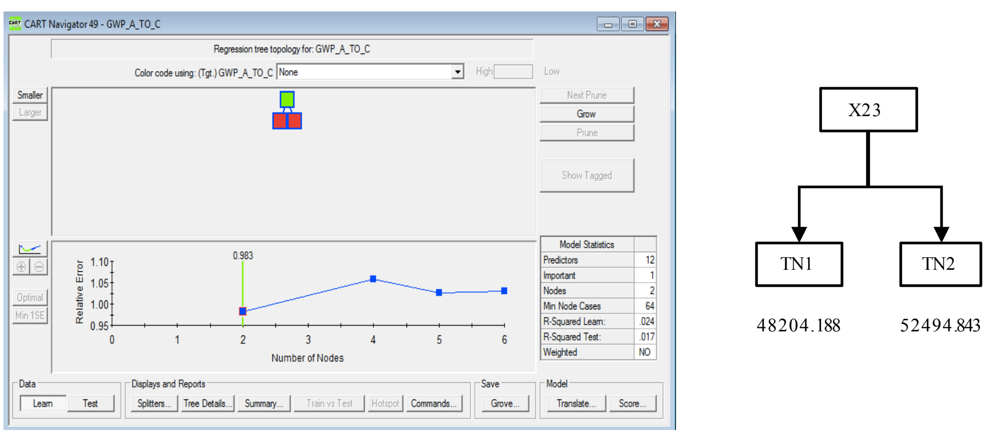
(a) (b)

Figure 5.14 a) Tree details; b) Tree model for “GWP” when all of the predictor variables are considered and the limit of splitting the nodes is “not split if there is less than 51 cases” based on Kung’s design



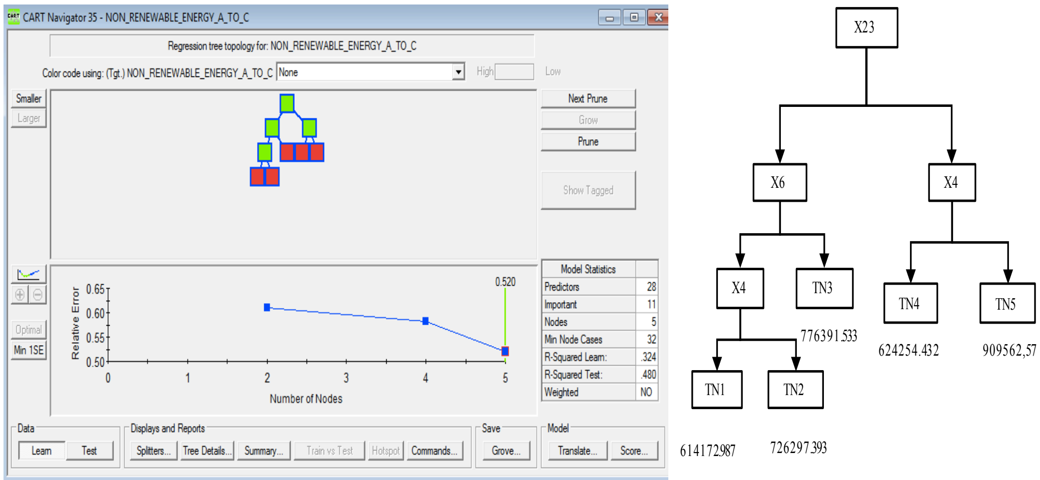
(a) (b)

Figure 5.15 a) Tree details; b) Tree model for “GWP” when all of the predictor variables are considered and the limit of splitting the nodes is “not split if there is less than 52 cases” based on Kung’s design



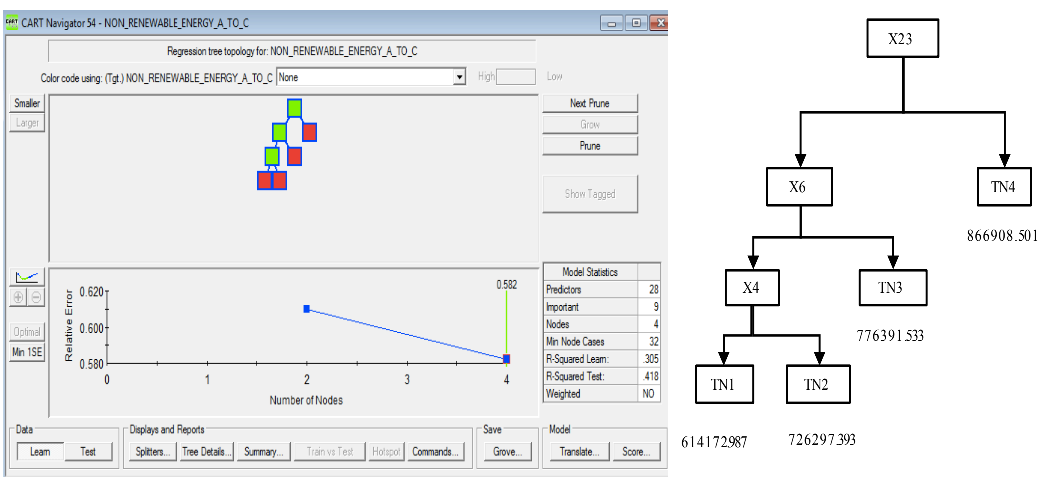
(a) (b)

Figure 5.16 a) Tree details; b) Tree model for “GWP” when only categorical predictor variables are considered and the limit of splitting the nodes is “not split if there is less than 60 cases” based on Kung’s design



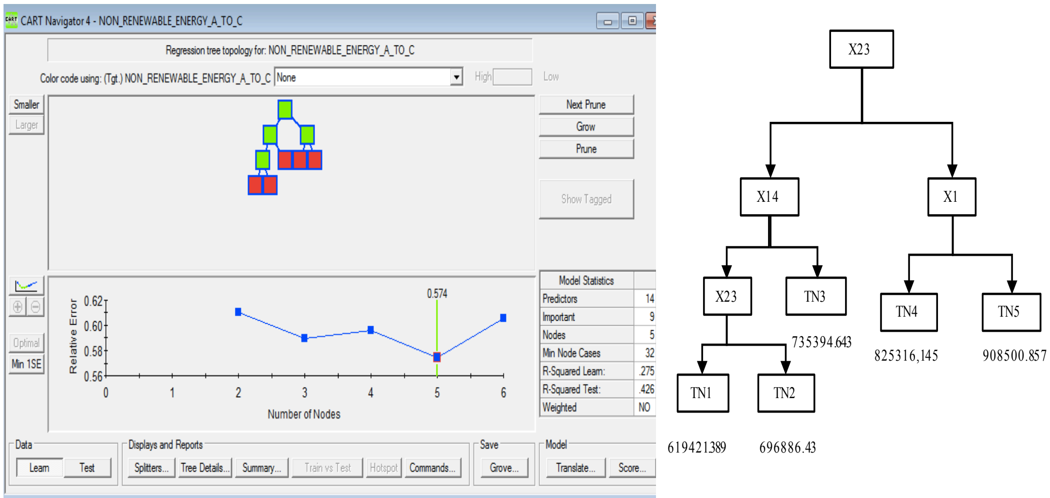
(a) (b)

Figure 5.17 a) Tree details; b) Tree model for “GWP” when all of the predictor variables are considered and the limit of splitting the nodes is “not split if there is less than 64 cases” based on Kung’s design



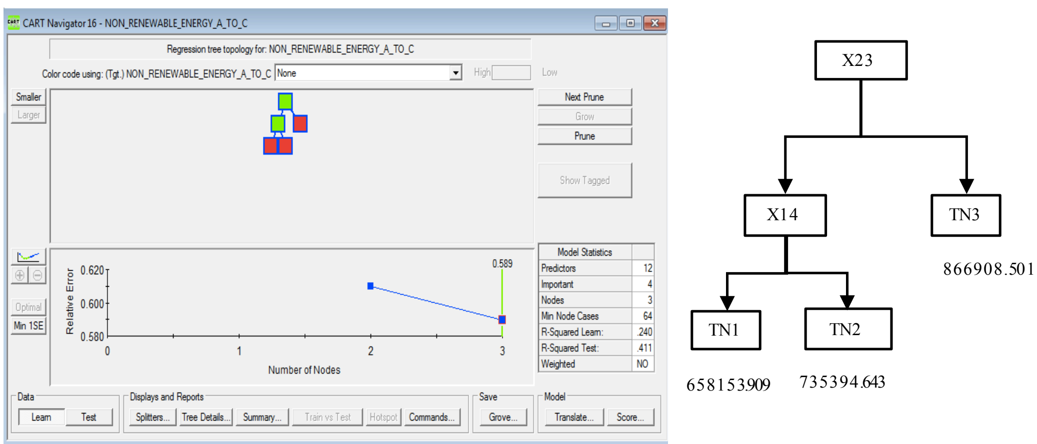
(a) (b)

Figure 5.18 a) Tree details; b) Tree model for “non-renewable energy” when all of the predictor variables are considered and the limit of splitting the nodes is “not split if there is less than 65 cases” based on Kung’s design



(a) (b)

Figure 5.19 a) Tree details; b) Tree model for “non-renewable energy” when only categorical predictor variables are considered and the limit of splitting the nodes is “not split if there is less than 64 cases” based on Kung’s design



(a) (b)

Figure 5.20 a) Tree details; b) Tree model for “non-renewable energy” when only categorical predictor variables are considered and the limit of splitting the nodes is “not split if there is less than 65 cases” based on Kung’s design

As it can be seen in Table 5.12, in some situations, when all of the predictor variables are considered in the tree, some numerical variables show up as an important variable in the tree. For example, for “GWP,” when the limit to split the node is 52 cases, “Ground Floor

Construction-X4” and “Concrete slab on grade-X5,” which are numerical variables, show up in the tree model. This suggests that the interaction of these two numerical variables is important in the regression model. Thus, another situation is added to the situations to be investigated in this study; for the “GWP” and considering only the categorical variables to generate the tree, the interaction term between “Ground Floor Construction-X4” and “Concrete slab on grade-X5,” i.e., X4X5, is added to the predictor variables. In order to avoid any multi-colinearity between the variables X4 and X5, and their interaction term, it is required to standardize the interaction term. Thus, at first each variable should be standardized, and, then, the product of these variables will produce the standardized interaction term, stdX4X5. Is it important to note that in order to standardize a variable, it is required to, first, center the mean to zero, and then scale its variance to one.

Table 5-12 Summary of CART results for each case of investigation for Kung's design

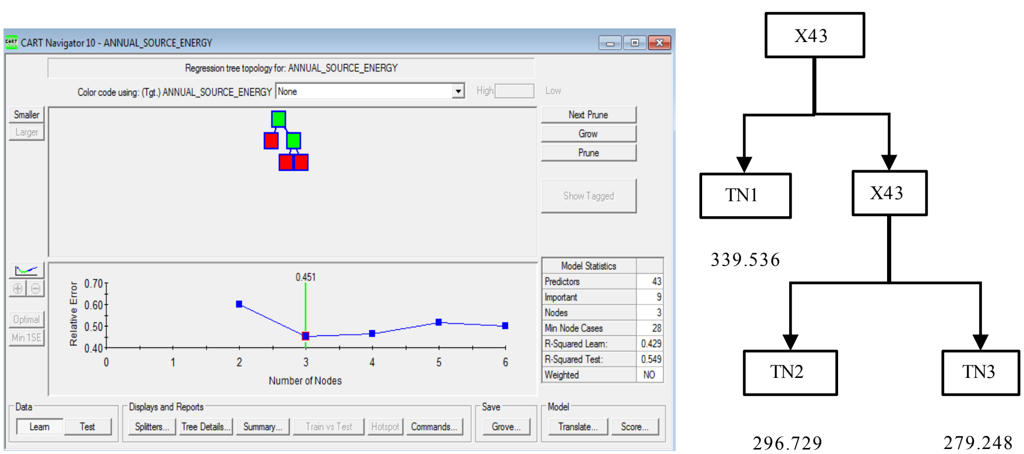
Response	Variables type in tree generation	Limits to split the node	Important variables	Node# 1	Node# 2	Node# 3	Node# 4	Node# 5
Annual source energy	All	67 Relative error = 0.489	X43	<=22.06	>22.06	>22.06	>22.06	>22.06
			X1	---	b	b	a	a
			X2	---	---	---	a, b	c, d
			X43	---	<=64.52	>64.52	---	---
			mean	342.97	290.182	281.886	289.368	301
			observation	33	44	35	38	42
		68 Relative error = 0.462	X43	<=24.44	>24.44	>24.44	>24.44	>24.44
			X1	---	b	b	a	a
			X2	---	---	---	a, b	c, d
			X43	---	<=64.52	>64.52	---	---
			mean	338.676	290.256	281.886	289.378	300.125
			observation	37	43	35	37	40
GWP	All	51 Relative error = 0.774	X4	<=6	>6	---	---	---
			mean	46285.001	55844.248	---	---	---
			observation	96	96	---	---	---
		52 Relative error = 0.803	X4	<=6	>6	>6	---	---
			X5	---	<=4500	>4500	---	---
			mean	46285.001	52532.361	59156.135	---	---
			observation	96	48	48	---	---
	Categorical	60	X23	c	a, b	---	---	---

		Relative error = 0.983	mean	48204.188	52494.843	---	---	---
			observation	64	128	---	---	---
Non-renewable energy	All	64 relative error= 0.695	X23	a, b	a, b	a, b	c	c
			X6	<=24.5	<=24.5	>24.5	---	---
			X4	<=6	>6	---	<=6	>6
			mean	614172.987	726297.393	776391.533	824254.432	909562.57
			observation	48	48	32	32	32
		65 Relative error = 0.756	X23	a, b	a, b	a, b	c	---
			X6	<=24.5	<=24.5	>24.5	---	---
			X4	<=6	>6	---	---	---
			mean	614172.987	726297.393	776391.533	866908.501	---
			observation	48	48	32	64	---
	Categorical	64 Relative error = 0.724	X23	a, b	a, b	a, b	c	c
			X14	a	a	b	---	---
			X23	a	b	---	---	---
			X1	---	---	---	b	a
			mean	619421.389	696886.43	735394.643	825316.145	908500.857
			observation	32	32	64	32	32
		65 Relative error = 0.760	X23	a, b	a, b	c	---	---
			X14	a	b	---	---	---
			mean	658153.909	735394.643	866908.501	---	---
			observation	64	64	64	---	---

In addition, when all of the predictors are considered to generate the tree for "non-renewable energy," among all of the numerical predictors, only "Ground Floor Construction-X4" and "Ground Floor Exterior/Cav insulation-X6" show up in the tree model. Here, the method explained in the previous paragraph is used to standardize the interaction term between X4 and X6. Ultimately, stdX4X6 is added to the numerical predictor variables that are used to predict the "non-renewable energy" when the tree is based only on the categorical variables. By adding these options to the previous situations under the investigation, the number of situations under the investigations becomes twelve situations.

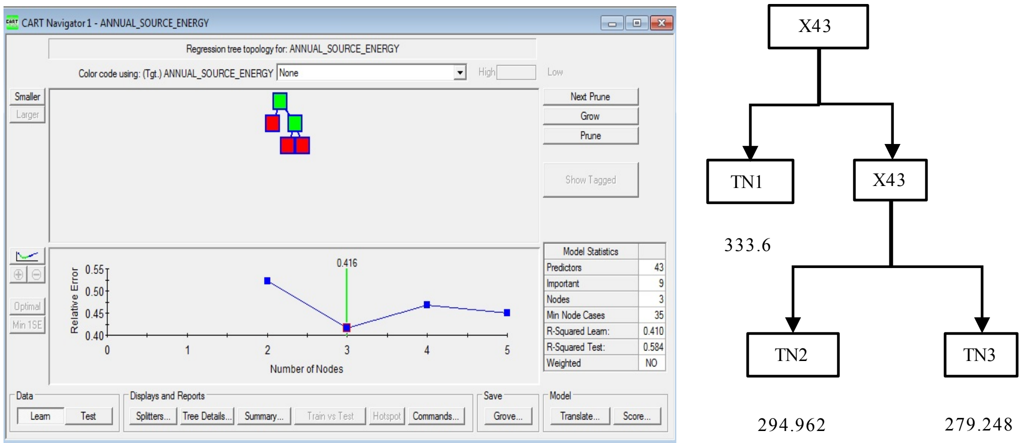
5.2.2 Martinez's Design

The trees generated in the CART module based on the selected thresholds for each of the above-mentioned cases are shown in Figure 5.21 to 5.30. Table 5.13 summarizes the results obtained from the CART module regarding the tree part by considering all of the predictors to generate the trees or by considering categorical variables only.



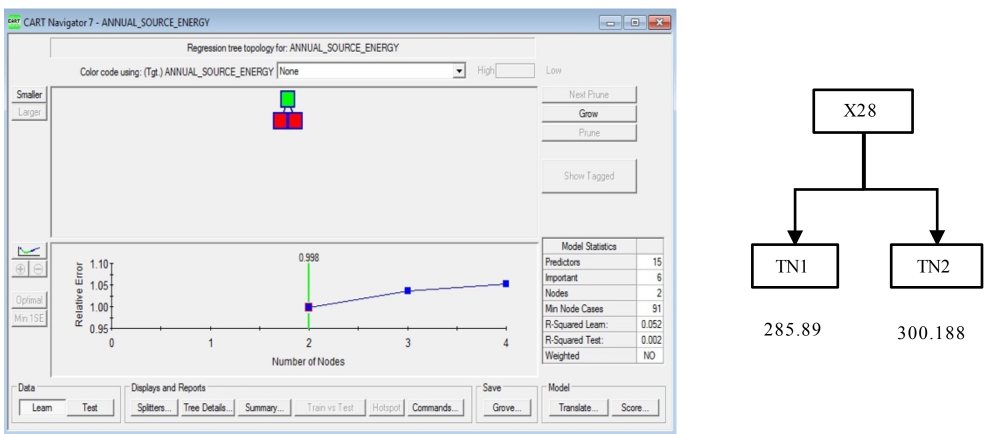
(a) (b)

Figure 5.21 a) Tree details; b) Tree model for "annual source energy" when all of the predictor variables are considered and the limit of splitting the nodes is "not split if there is less than 57 cases" based on Martinez's design



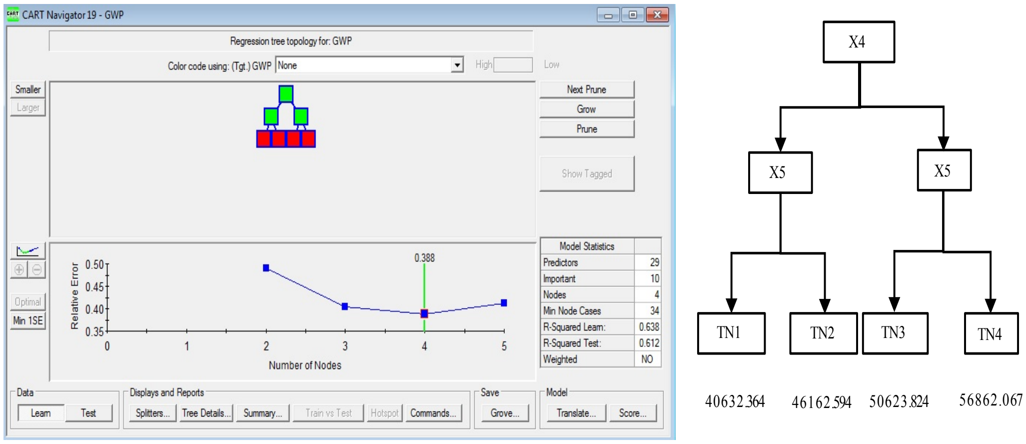
(a) (b)

Figure 5.22 a) Tree details; b) Tree model for "annual source energy" when all of the predictor variables are considered and the limit of splitting the nodes is "not split if there is less than 58 cases" based on Martinez's design



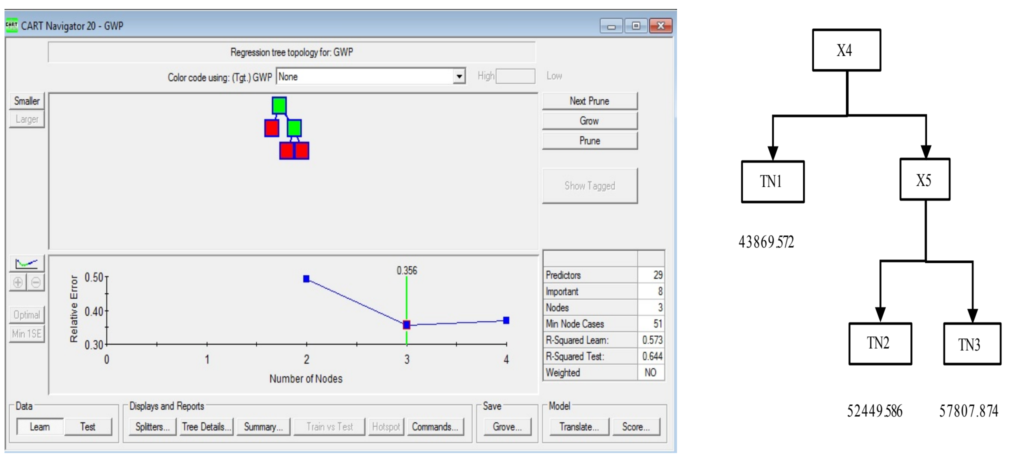
(a) (b)

Figure 5.23 a) Tree details; b) Tree model for "annual source energy" when only the categorical variables are considered and the limit of splitting the nodes is "not split if there is less than 60" based on Martinez's design



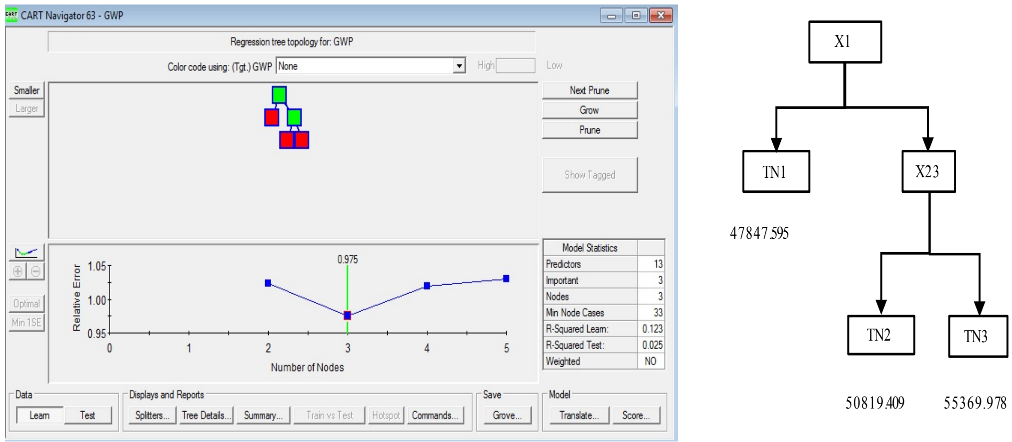
(a) (b)

Figure 5.24 a) Tree details; b) Tree model for "GWP" when all of the predictor variables are considered and the limit of splitting the nodes is "not split if there is less than 69" based on Martinez's design



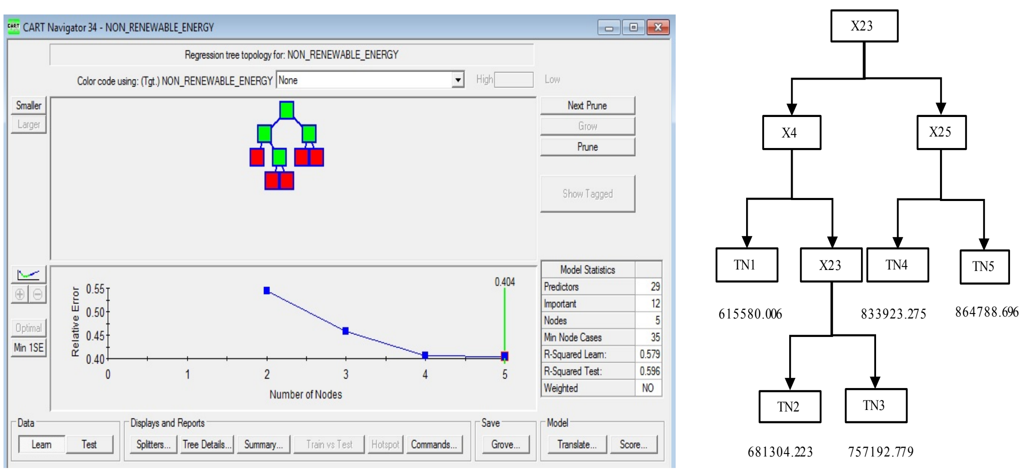
(a) (b)

Figure 5.25 a) Tree details; b) Tree model for "GWP" when all of the predictor variables are considered and the limit of splitting the nodes is "not split if there is less than 70 cases" based on Martinez's design



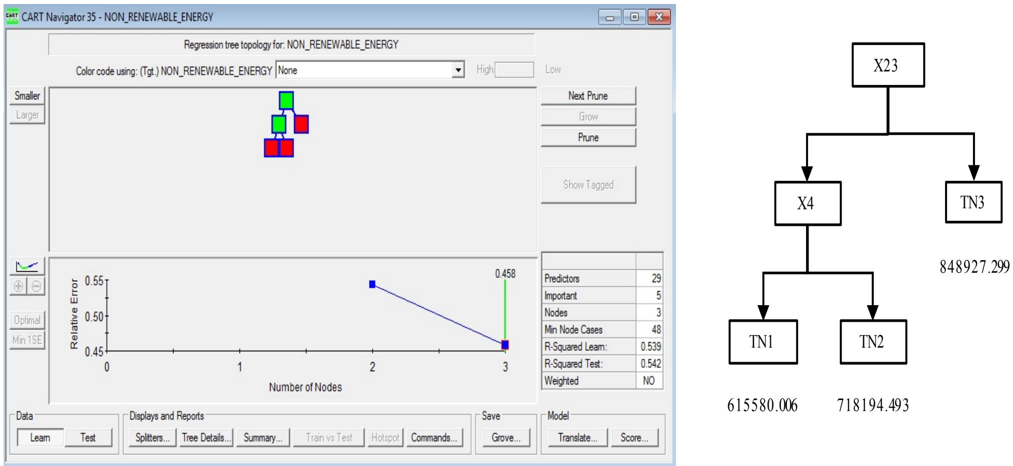
(a) (b)

Figure 5.26 a) Tree details; b) Tree model for “GWP” when only the categorical variables are considered and the limit of splitting the nodes is “not split if there is less than 60 cases” based on Martinez’s design



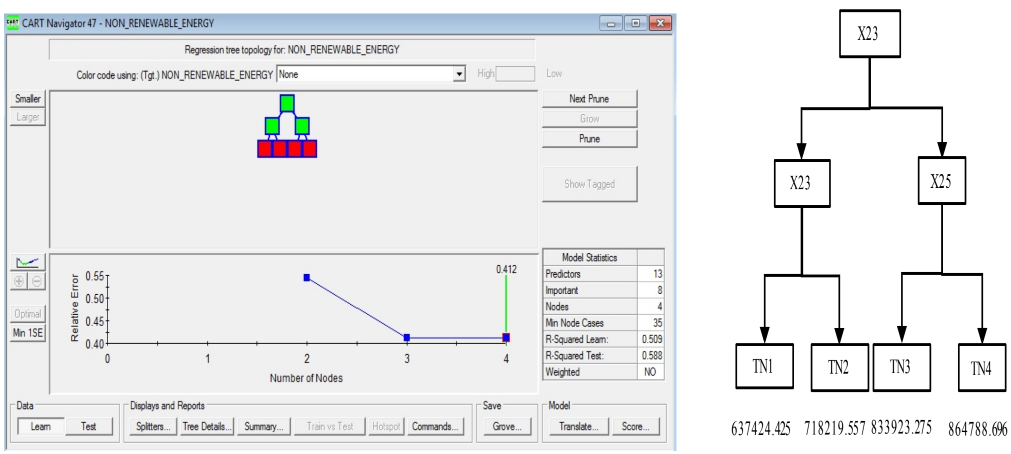
(a) (b)

Figure 5.27 a) Tree details; b) Tree model for “non-renewable energy” when all of the predictor variables are considered and the limit of splitting the nodes is “not split if there is less than 72 cases” based on Martinez’s design



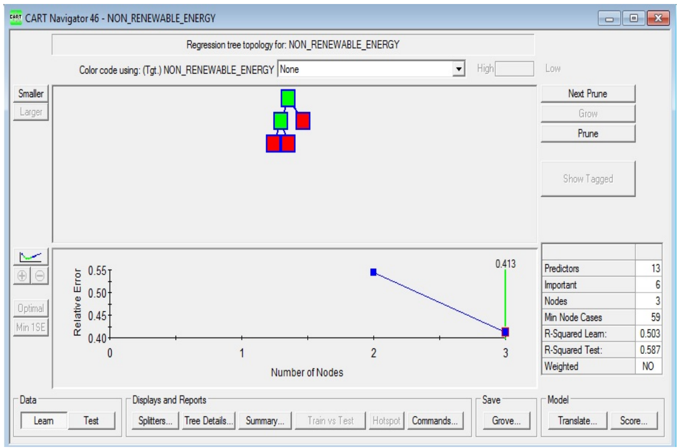
(a) (b)

Figure 5.28 a) Tree details; b) Tree model for “non-renewable energy” when all of the predictor variables are considered and the limit of splitting the nodes is “not split if there is less than 73 cases” based on Martinez’s design

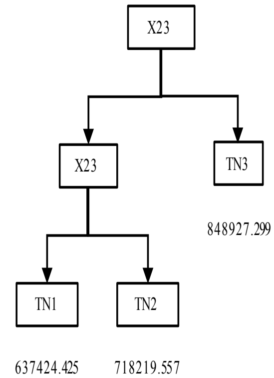


(a) (b)

Figure 5.29 a) Tree details; b) Tree model for “non-renewable energy” when only the categorical variables are considered and the limit of splitting the nodes is “not split if there is less than 72 cases” based on Martinez’s design



(a)



(b)

Figure 5.30 a) Tree details; b) Tree model for “non-renewable energy” when only the categorical variables are considered and the limit of splitting the nodes is “not split if there is less than 73 cases” based on Martinez’s design

Table 5-13 Summary of CART results for each case of investigation for Martinez's design

response	variables type in tree generation	limits to split the node	important variables	Node# 1	Node# 2	Node# 3	Node# 4	Node# 5
1	All	57	X43	<=18.95	>18.95 & <=47.96	>47.96	---	---
			mean	339.536	296.729	279.248	---	---
			observation	28	59	105	---	---
		58	X43	<=22.74	>22.74 & <=47.96	>47.96	---	---
			mean	333.6	294.962	279.248	---	---
			observation	35	52	105	---	---
	Categorical	60	X28	a	b	---	---	---
			mean	285.89	300.188	---	---	---
			observation	91	101	---	---	---
2	All	69	X4	<=6	<=6	>6	>6	---
			X5	<=4500	>4500	<=3500	>3500	---
			mean	40632.264	46162.594	50623.824	56862.067	---
			observation	34	48	34	76	---
		70	X4	<=6	>6	>6	---	---
			X5	---	<=4500	>4500	---	---
			mean	43869.572	52449.586	57807.874	---	---
			observation	82	59	51	---	---

	Categorical	60	X1	b	a	a	---	---
			X23	---	a, c	b	---	---
			mean	47847.595	50819.409	55369.978	---	---
			observation	90	69	33	---	---
3	All	72	X23	a, b	a, b	a, b	c	c
			X4	<=6	>6	>6	---	---
			X25	---	---	---	b	a
			X23	---	a	b	---	---
			mean	615580.006	681304.223	757192.779	101230.145	864788.696
			observation	48	37	35	37	35
		73	X23	a, b	a, b	c	---	---
			X4	<=6	>6	---	---	---
			mean	615580.006	718194.493	848927.299	---	---
			observation	48	72	72	---	---
	Categorical	72	X23	a	b	c	c	---
			X25	---	---	b	a	
			mean	637424.43	718219.557	833923.275	864788.696	---
			observation	61	59	37	35	---
		73	X23	a	b	c	---	---
			mean	637424.425	718219.557	848927.299	---	---

			observatio n	61	59	72	---	---

5.3 Fitting the Regression models

5.3.1 Kung's Design

In this step, it is required to fit the regression line on each response variable separately based on the generated trees, shown in Table 5.12. Since in this study there is a large number of predictor variables, the stepwise method is used to fit the regression line on each response variable separately ($\alpha = 0.05$). For example, for "annual source energy," when the tree is generated based on all predictors and the limit to split the node is 67 cases, the tree has five TNs. Thus, Statistical Analysis Software (SAS) is used to regress the numerical variables on the response 1, "annual source energy." The fitted models for Kung's design based on the treed regression approach are shown in Table 0.4-0.15 (Appendix C).

Table 5-14 Important variables for Kung's design based on tree and regression models

Response	Type of variables involved in the tree	Limits for splitting the tree	Important variables
Annual source energy	All	67 Relative error = 0.489	Tree: **X1, X2, X43** Fitted Regression Line without Interaction: **X6, X9 , X10, X21, X36, X37, X38, X43, X44**
		68 Relative error = 0.462	Tree: **X1, X2, X43** Fitted Regression Line without Interaction: **X6, X9 , X10, X21, X36, X37, X38, X43, X44**
GWP	Categorical	60 Relative error = 0.983	Tree: **X23** Fitted Regression Line without Interaction: **X3, X4, X5, X6, X19, X20, X27, X35, X37** Fitted Regression Line with Interaction: **Same as model without interaction**
	All	51 Relative error = 0.774	Tree: **X4** Fitted Regression Line without Interaction: **X5, X6, X19, X20, X27, X35**
		52	Tree: **X4, X5**

		Relative error = 0.803	Fitted Regression Line without Interaction: **X5, X6, X8, X10, X19, X20, X31, X35, X36, X38**
Non-renewable energy	Categorical	64 Relative error = 0.724	Tree: **X1, X14, X23** Fitted Regression Line without Interaction: **X4, X5, X6, X8, X9, X10, X18, X19, X20, X27, X31, X35, X36, X37** Fitted Regression Line with Interaction: **X4, X5, X6, X8, X9, X10, X18, X19, X20, X27, X31, X35, X36, stX4X6**
		65 Relative error = 0.760	Tree: **X14, X23** Fitted Regression Line without Interaction: **X3, X4, X5, X6, X9, X10, X19, X20, X27** Fitted Regression Line with Interaction: **X3, X4, X5, X6, X9, X10, X19, X20, X27, X35, X38, stX4X6**
	All	64 Relative error = 0.695	Tree: **X4, X6, X23** Fitted Regression Line without Interaction: **X3, X5, X6, X10, X19, X20, X27, X35, X37**
		65 Relative error = 0.695	Tree: **X4, X6, X23** Fitted Regression Line without Interaction: **X3, X4, X5, X6, X10, X19, X20, X27, X35**

The predictor variables that affect the performance metrics are summarized in two different formats in Table 5.14 and 5.15. These tables can help in determining the predictor variables that play significant roles on the response variables. For example, "Ground Floor Exterior/Cav insulation-X6" is the only predictor variable that shows up to be significant regardless of the studied response variable. In addition, it is concluded from these tables that "Max occupancy – Dining area-X43," "Foot Print Shape-X1," and "Orientation-X2" are important only for "annual source energy," when all variables are considered in the tree generation step. Thus, it seems that eQUEST performance metric are affected by the siting options, the foundation system, the wall system, the roof system, the window system, and the ventilation system. The "GWP," when the tree is based on all predictors is affected by the foundation system, the wall system, the roof system, and the window system. But, the "GWP," when the tree is based on only categorical variables, is affected by the foundation system, the room system, and the window system. The "non-renewable energy" when the tree is based on all variables is impacted by the foundation system, the wall system, the roof system, and the window system, while the "non-renewable energy" when the tree is based on only categorical variables is affected by the siting options, the foundation system, the wall system, the roof system, and the window system.

Even when the interaction term is considered in the modeling of "GWP," the predictor variables affecting the performance metrics do not change. However, considering "non-renewable energy" when the bigger tree is considered, the only change is removing "Total Window Area % East- X37" and adding the interaction term, stdX4X6. This interaction term represents the interaction between "Ground Floor Construction" and "Ground Floor Exterior/Cav insulation," which are both from the stage of floor system. In addition, regarding the smaller tree, by considering the interaction stdX4X6, only "Total Window Area % North-X35," "Total Window Area % West-X38," and "stdX4X6" are added to the effective variables.

Table 5-15 Summary of important variables in treed regression method for Kung's design based on variable categories

Variable category	Important variable	Type of variable	All of variables in tree			Only categorical variables in tree	
			ASE	GWP	NON-RNE	GWP	NON-RNE
Siting options	X1	Ctg.	*	---	---	---	Big tree
	X2	Ctg.	*	---	---	---	---
Foundation system	X3	Num.	---	---	*	*	Small tree
	X4	Num.	---	*	*	*	*
	X5	Num.	---	*	*	*	*
	X6	Num.	*	*	*	*	*
Wall system	X8	Num.	---	Big tree	---	---	Big tree
	X9	Num.	*	---	---	---	*
	X10	Num.	*	Big tree	*	---	*
	X14	Ctg.	---	---	---	---	*
Roof system	X18	Num.	---	---	---	---	Big tree
	X19	Num.	---	*	*	*	*
	X20	Num.	---	*	*	*	*
	X21	Num.	*	---	---	---	---
	X23	Ctg.	---	---	*	*	*
	X27	Num.	---	Small tree	*	*	*
Window system	X31	Num.	---	Big tree	---	---	Big tree
	X35	Num.	---	*	*	*	Big tree
	X36	Num.	*	Big tree	---	---	Big tree
	X37	Num.	*	---	Big tree	*	Big tree
	X38	Num.	*	Big tree	---	---	---
Ventilation system	X43	Cunt.	*	---	---	---	---
	X44	Cunt.	*	---	---	---	---
interaction	stdX4X6	---	---	---	---	---	*

5.3.2 *Martinez's Design*

In this step, the stepwise regression method is used to fit the regression line on each response variable separately based on the generated trees shown in Table 5.16 ($\alpha = 0.05$). for Martinez's design, as it can be seen in all of the predictors that show up in the trees are categorical, except X4 and X5. It seems that X4 is important as a numerical predictor for "GWP" and "non-renewable energy," and X5 is a numerical predictor only for the tree generation for "non-renewable energy." Thus, the interaction between X4 and X5 is the only case that is considered in fitting the regression line on the "non-renewable energy." Table 5.17 provides the results of the treed regression approach for the outputs based on Martinez's design. The fitted models for Martinez's design based on the treed regression approach are shown in Table 0.32-0.42 (Appendix C).

Table 5-16 Important variables for Martinez's design based on tree and regression models

Response	Type of variables involved in the tree	Limits for splitting the tree	Important variables
Annual source energy	Categorical	60 Relative error =0.998	Tree: **X28** Fitted Regression Line without Interaction: **X10, X35, X36, X37, X38, X43, X44**
	All	57 Relative error =0.451	Tree: **X43** Fitted Regression Line without Interaction: **X4, X6, X10, X35, X36, X37, X38, X43, X44, X48, X52**
		58 Relative error =0.416	Tree: **X43** Fitted Regression Line without Interaction: **X4, X6, X10, X35, X36, X37, X38, X42, X43, X44, X48**
GWP	Categorical	60 Relative error = 0.975	Tree: **X1, X23** Fitted Regression Line without Interaction: **X4, X5, X6, X10, X18, X19, X20, X35, X36, X37, X38** Fitted Regression Line with Interaction:

			X4, stdX4X5, X5, X6, X10, X18, X19, X20, X35, X36, X37, X38
	All	69 Relative error = 0.388	Tree: **X4, X5** Fitted Regression Line without Interaction: **X3, X5, X6, X10, X19, X20, X27, X38**
		70 Relative error = 0.356	Tree: **X4, X5** Fitted Regression Line without Interaction: **X3, X5, X6, X9, X10, X19, X20, X35, X38**
Non-renewable energy	Categorical	72 Relative error = 0.412	Tree: **X23, X25** Fitted Regression Line without Interaction: **X3, X4, X5, X6, X8, X9, X10, X19, X20, X35, X36**
		73 Relative error = 0.413	Tree: **X23** Fitted Regression Line without Interaction: **X3, X4, X5, X6, X8, X10, X19, X20, X27, X35, X36**
	All	72 Relative error = 0.404	Tree: **X4, X23, X25** Fitted Regression Line without Interaction: **X3, X4, X5, X6, X9, X10, X19, X20, X35, X36**
		73 Relative error = 0.458	Tree: **X4, X23** Fitted Regression Line without Interaction: **X3, X4, X5, X6, X10, X19, X20, X27, X36**

Tables 5.16 and 5.17 show the important predictor variables based on the performance metrics from either eQUEST or ATHENA. These tables are similar in content, but different in format. As it can be seen, “Walls Additional insulation-X10,” which is the wall system, and “Total Window Area % North-X35,” which is the window system are the predictor variables that show up to be significant for either of performance metrics from eQUEST and ATHENA. However, X35 affects “GWP,” when the tree generation is based on all of the predictors in the smaller tree, and it effects “non-renewable energy” when tree generation is based on all of the predictors in the bigger tree. It seems that “Foot Print Shape-X1” from siting options, and “Roof Load Bearing-X18” from roof system are only

important for “GWP,” when the tree is generated based on only the categorical predictor variables. “Ventilation – living space-X42,” “Ventilation – Dining area-X44,” “ventilation – Corridor -X48,” and “ventilation -All others-X52” from ventilation system, and “Max occupancy – Dining area-X43” from max occupancy design only affect performance metrics from eQUEST (“annual source energy”) when the tree is generated by considering all of predictors. However, “Max occupancy – Dining area-X43” and “Ventilation – Dining area-X44” affect “annual source energy,” regardless of the type of the variables being considered in the tree generation. “Window Type-X28” from window system is important only for “annual source energy,” considering categorical predictors only in the tree generation. “Ground Floor Construction-X4” and “Ground Floor Exterior/Cav insulation-X6” from foundation system, are important variables for all of the performance metrics regardless of the type of the variable being considered to generate the tree, except for “annual source energy.” These predictors are important for “annual source energy” only when the tree generation is based on all of the predictors.

Also, “Total Window Area % West-X38” from the wall system affects “annual source energy” and “GWP” only. “Roof Exterior insulation-X19” and “Roof Additional insulation-X20” from the roof system, and “Concrete slab on grade-X5” from the foundation system are important predictors only for responses from ATHENA. This is also true for “Ground Floor Interior insulation-X3” from the foundation system, and “Ceiling Exterior finish-X23” from the roof system, however, there are some exceptions. In other word, X3 is not important only for “GWP” when only categorical predictors are considered in the tree part, and X23 is not important for “GWP” only when the tree is generated based on all of the variables. The predictor “Roof Type-X25” is only important for the non-renewable energy regardless of the type of variable used in the tree generation step. However, this predictor is only important for the biggest trees, when either only categorical variables are considered in tree generation or all of the variables are considered. In addition, it seems that “Walls Interior Insulation-X8” from the wall system is important only for “non-renewable energy” when the tree generation is based on only the categorical variables. “Total Window Area % South-X36,” and “Total Window Area % East-X37” from the window system shows up for “annual source energy” in all of the models, and for “GWP,” when the tree is based on the categorical variables only. In addition, X36 seems to be important for “non-renewable energy” in all of trees. Also, “Walls Exterior insulation-X9” seems to affect the biggest tree of “non-renewable energy” regardless of the type of variables, and the smallest tree of “GWP” based on all of the predictors. This is true about “Roof Decking thickness-X27,” but considering the biggest tree for “GWP” and smallest trees for “non-renewable energy.”

Table 5-17 Summary of important variables in treed regression method for Martinez's design based on variable categories

Variable category	Important variable	All of variables in tree			Only categorical variables		
		ASE	GWP	NON-RNE	ASE	GWP	NON-RNE
Siting options	X1	---	---	---	---	*	---
Foundation system	X3	---	*	*	---	---	*
	X4	*	*	*	---	*	*
	X5	---	*	*	---	*	*
	X6	*	*	*	---	*	*
Wall system	X8	---	---	---	---	---	*
	X9	---	small tree	Big tree	---	---	Big tree
	X10	*	*	*	*	*	*
Roof system	X18	---	---	---	---	*	---
	X19	---	*	*	---	*	*
	X20	---	*	*	---	*	*
	X23	---	---	*	---	*	*
	X25	---	---	Big tree	---	---	Big tree
	X27	---	Big tree	Small tree	---	---	Small tree
Window system	X28	---	---	---	*	---	---
	X32	---	---	---	---	---	---
	X35	*	Small tree	Big tree	*	*	*
	X36	*	---	*	*	*	*
	X37	*	---	---	*	*	---
	X38	*	*	---	*	*	---
Ventilation system	X42	*	---	---	---	---	---
	X43	*	---	---	*	---	---
	X44	*	---	---	*	---	---
	X48	*	---	---	---	---	---
	X52	*	---	---	---	---	---

Thus, it seems that response 1 ("annual source energy") and response 2 ("global warming potential"), regardless of the variables that show up into the tree, are affected by the siting option, the foundation system, the wall system, the roof system, and the window system, based on the second design. "Non-renewable energy" is affected by the foundation system, the wall system, the room system, and the window system, based on Martinez's design.

5.4 Fitting MARS Model

5.4.1 Kung's Design

Some of the parameter settings in the MARS method can take several values and may change the fitted model. These parameters include the maximum number of basis functions (MBF), the degree of freedom of the interaction terms, which represents the maximum number of variables that are in the interaction terms or maximum interaction (it is represented as the maximum interaction (MI) in Salford Systems MARS software), and the minimum number of observations between the knots (MOBN). In this study, several MBFs were selected to be investigated: 20, 30, 50, 100, and 150. It is seen (see Table 0.2 in Appendix B) that the test performance R-square does not change when the MBF value changes from 20 into 30. In addition, when the MBF value changes from 100 to 150, the number of actual basis functions (=38) does not change, and the test performance R-square is the same for the MBF values of 100 and 150. Thus, the MBF values of 30 and more than 100 are not investigated, and it is decided to consider three values, i.e., 20, 50, and 100 for MBF.

Also, For MI three value; in terms of 1 (no interaction term), 2 (two-factor interaction term), and 3 (three-factor interaction term) are considered. In addition, for MOBN three values are assumed, 2, 5, and 10. Since selecting only one model as the best model among all of the models (see Appendix B) is not possible, it was decided to select 20 percent of the models as the best ones, based on the lowest predicted residual error sum of squares

(PRESS) value and mean of absolute relative error (MARE), which is discussed in the validation of MARS (section 6.2). Thus, as it can be seen in Table 5.18, which shows the important variables based on the MARS method for Kung's design, only some of the values for MBF, MI, and MOBN show up as the setup of the better models. The validation data set with 96 runs, which was discussed in Section 4.1, is used for testing. The fitted models for Kung's design based on the MARS approach are shown in Tables 0.16-0.31 (Appendix C).

Table 5-18 Summary of important variables in the MARS method for Kung's design based on variable categories

variable category	important variable		R1										R2			R3		
		MI	3	3	3	2	2	2	2	2	2	3	1	1	2	1	1	3
		MOBN	2	2	2	2	2	2	5	5	5	5	2	2	10	2	2	5
		MBF	20	50	100	50	20	100	50	100	20	20	20	50	50	20	50	100
siting options	X1		*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	---
	X2		*	*	*	*	*	*	*	*	*	*	*	*	*	*	---	---
Foundation system	X3		---	---	---	*	*	*	---	---	---	---	*	*		*	*	
	X4		---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
	X5		---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
	X6		*	*	*	*	*	*	*	*	*	*	---	---	---	---	---	---
	X7		---	---	---	*	---	---	---	---	---	---	---	---	---	---	---	---
Wall system	X9		---	---	---	*	*	---	*	*	---	---	---	---	---	---	---	---
	X10		*	*	*	*	*	*	*	*	*	*	---	---	---	---	---	---
	X14		---	---	---	---	---	---	---	---	---	---	---	---	---	*	*	---
	X15		---	---	---	*	---	---	---	---	---	---	---	---	---	---	---	---
Roof system	X19		---	---	---	---	---	---	---	---	---	---	---	---	---	*	---	---
	X20		---	---	---	---	---	---	---	---	---	---	*	*	*	*	*	---
	X23		---	---	---	---	---	---	---	---	---	---	*	---	---	*	*	*
	X24		---	---	---	*	*	*	---	---	---	---	---	---	---	---	---	---
	X27		---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	*
Window system	X29		---	---	---	---	---	---	---	---	---	---	---	---	*	*	---	---
	X32		---	---	---	---	---	---	*	*	*	*	---	---	---	---	---	---
	X33		---	*	*	*	---	---	---	---	---	---	---	---	---	---	---	---
	X35		*	*	*	*	---	---	---	---	---	---	---	---	---	*	*	---
	X36		*	---	---	*	---	---	---	---	---	---	---	---	*	---	---	---
	X37		---	---	---	*	---	---	*	*	*	*	---	---	---	---	---	---

	X38	*	*	*	*	*	*	---	---	*	*	---	---	---	*	---	---
Ventilation system	X43	*	*	*	*	*	*	*	*	*	*	---	---	---	---	---	---
	X44	*	*	*	*	*	*	*	*	*	*	---	---	---	---	---	---
	X52	---	---	---	---	---	---	*	---	---	---	---	---	---	---	---	---
2-factor interaction	X1X2	*	*	*	*	*	*	*	*	*	*	---	---	---	---	---	---
	X1X6	*	*	*	*	*	---	*	*	---	---	---	---	---	---	---	---
	X1X9	---	---	---	---	---	---	*	*	---	---	---	---	---	---	---	---
	X1X15	---	---	---	*	---	---	---	---	---	---	---	---	---	---	---	---
	X1X20	---	---	---	---	---	---	*	---	---	---	---	---	---	---	---	---
	X1X23	---	---	---	---	---	---	*	---	---	---	---	---	---	---	---	---
	X1X32	---	---	---	---	---	---	*	*	*	*	---	---	---	---	---	---
	X1X35	---	---	---	---	---	---	*	---	---	---	---	---	---	---	---	---
	X1X37	---	---	---	---	---	---	*	---	*	*	---	---	---	---	---	---
	X1X43	---	---	---	---	---	---	*	---	---	---	---	---	---	---	---	---
	X1X52	---	---	---	---	---	---	*	---	---	---	---	---	---	---	---	---
	X2X38	*	---	---	*	---	---	*	*	*	*	---	---	---	---	---	---
	X3X6	---	---	---	*	*	*	*	*	---	---	---	---	---	---	---	---
	X4X38	---	---	---	---	---	---	*	---	---	---	---	---	---	---	---	---
	X5X6	---	---	---	---	---	---	---	---	---	---	---	---	*	---	---	---
	X6X20	---	---	---	---	---	---	---	---	---	---	---	---	*	---	---	---
	X6X36	---	---	---	---	---	---	---	---	---	---	---	---	*	---	---	---
	X6X40	---	---	---	---	---	---	*	---	---	---	---	---	---	---	---	---
	X7X43	---	---	---	*	---	---	---	---	---	---	---	---	---	---	---	---
	X9X38	---	---	---	*	---	---	---	---	---	---	---	---	---	---	---	---
	X10X24	---	---	---	*	*	---	---	---	---	---	---	---	---	---	---	---
	X10X33	---	---	---	*	---	---	---	---	---	---	---	---	---	---	---	---
	X10X35	*	*	---	*	---	---	---	---	---	---	---	---	---	---	---	---
	X10X37	---	---	---	---	---	---	*	*	---	---	---	---	---	---	---	---

	X10X43		---	---	*	---	---	---	*	---	---	---	---	---	---	---	---	---
	X15X41		---	---	---	---	---	---	*	---	---	---	---	---	---	---	---	---
	X20X41		---	---	---	---	---	---	*	---	---	---	---	---	---	---	---	---
	X23X27		---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	*
	X29X35		---	---	---	---	---	---	---	---	---	---	---	---	*	---	---	---
	X36X38		---	---	---	*	---	---	---	---	---	---	---	---	---	---	---	---
	X37X43		---	---	---	*	---	---	---	---	---	---	---	---	---	---	---	---
	X38X42		---	---	---	---	---	---	*	---	---	---	---	---	---	---	---	---
	X42X44		---	---	---	---	---	---	*	---	---	---	---	---	---	---	---	---
	X43X44		*	*	---	*	*	*	*	*	*	*	---	---	---	---	---	---
3-factor interaction	X1X2X38		---	*	*	---	---	---	---	---	---	---	---	---	---	---	---	---
	X1X9X33		---	*	*	---	---	---	---	---	---	---	---	---	---	---	---	---

Based on Table 5.18, it seems that the categories of variables that affect the "annual source energy" (eQUEST output) are the siting options, the foundation system, the wall system, the roof system, the window system, and the ventilation system. However, the roof system affects the eQUEST performance metrics only when either of MI and MOBN takes value of 2. The categories of variables that impact the "GWP" from ATHENA, are the siting option, the foundation system, the roof system, and the window system. In addition, the siting option, the foundation system, the wall system, the roof system, and the window system affect the "non-renewable energy" from ATHENA.

5.4.2 Martinez's Design

In this step, the MARS method is used to analyze the performance of the design based on Martinez's design. The same values for MBF, MI, and MOBN used for MARS method in Kung's design are used here; for MI, 1, 2, and 3 maximum interaction, and 50, and 100 values for MBF, and for MOBN, three values of 2, 5, and 10 are used. Similar to Kung's design, only some of these values show up as the setup of the better models for Martinez's design. However, the values of the better model for Martinez's design might be different from the values in Kung's design. Table 5.19 shows the important variables based on MARS method for Martinez's design. All of the assumptions in the MARS module were similar to the assumptions used in Kung's design. Here, the testing dataset with 96 runs of the simulation described in Section 4.3 for Martinez's design is used as the testing dataset in MARS.

The scenarios represented in Table 5.19 were selected based on the lowest PRESS and MARE values (which will be discussed in the validation of MARS in section 6.2) among all of the scenarios for the MARS approach based on Martinez's design (see Table 0.3 in Appendix B). The fitted models of these scenarios are shown in Table 0.43-0.61 (Appendix C).

Table 5-19 Summary of important variables in the MARS method for Martinez’s design based on variable categories

variable category	important variable		Annual source energy							GWP						Non-renewable energy					
		MI	2	3	1	2	2	2	1	2	1	1	3	2	1	2	1	1	1	3	3
		MOBN	10	10	10	10	2	5	5	10	2	2	2	10	2	2	2	2	2	5	5
		MBF	20	100	20	100	20	20	20	20	100	50	20	50	20	20	20	100	50	100	50
siting options	X1		*	---	*	---	---	---	*	---	---	---	---	---	---	---	---	---	---	---	---
	X2		*	---	*	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
Foundation system	X3		---	---	---	---	---	---	---	---	---	---	---	---	---	*	---	*	*	---	---
	X4		---	---	---	---	---	---	---	*	*	*	*	*	*	*	*	*	*	*	*
	X5		---	---	---	---	---	---	---	*	*	*	*	*	*	*	*	*	*	*	*
	X6		*	---	*	---	---	---	*	*	*	*	*	*	*	*	*	*	*	*	*
Wall system	X8		---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	*	---	---
	X9		---	---	---	---	---	---	---	---	*	*	---	---	*	---	---	---	*	---	---
	X10		---	---	---	---	---	---	---	*	*	*	*	*	*	*	*	*	*	*	---
	X11		---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	*	---	---
	X13		---	---	---	---	---	---	---	---	*	*	---	---	---	---	---	---	---	---	---

	X14	---	---	---	---	---	---	---	---	*	*	---	---	---	*	*	*	*	*	*
Roof system	X19	---	---	---	---	---	---	---	*	*	*	*	*	*	*	*	*	*	*	*
	X20	---	---	---	---	---	---	---	*	*	*	*	*	*	*	*	*	*	*	*
	X23	---	---	---	---	---	---	---	*	*	*	*	*	*	*	*	*	*	*	*
	X25	---	---	---	---	---	---	---	---	*	*	---	---	---	---	---	---	*	---	---
	X27	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	*	---	---
Window system	X28	---	---	---	---	---	---	---	*	*	---	*	*	---	*	---	*	*	*	---
	X29	---	---	---	---	---	---	---	---	*	*	---	---	---	---	---	---	*	---	---
	X32	---	---	---	---	---	---	---	*	*	*	*	*	*	*	*	*	*	---	---
	X35	---	---	*	---	---	---	---	*	*	*	*	*	*	*	---	---	*	---	---
	X36	*	---	*	---	---	---	*	*	*	---	*	*	---	*	*	*	---	*	---
	X37	*	---	*	---	---	---	*	---	---	---	---	---	---	---	---	---	*	---	---
	X38	*	---	*	---	---	---	*	---	*	*	---	---	---	---	---	---	*	---	---
Ventilation system	X40	*	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---

	X43		*	*	*	*	*	*	*	---	---	---	---	---	---	---	---	---	---	---	---
	X44		*	*	*	*	*	*	*	---	---	---	---	---	---	---	---	---	---	---	---
2-factor interaction	X1X2		*	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
	X1X40		*	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
	X3X23		---	---	---	---	---	---	---	---	---	---	---	---	---	*	---	---	---	---	---
	X4X5		---	---	---	---	---	---	---	*	---	---	---	*	---	---	---	---	---	---	---
	X4X28		---	---	---	---	---	---	---	---	---	---	---	---	---	*	---	---	---	*	---
	X20X35		---	---	---	---	---	---	---	---	---	---	---	*	---	---	---	---	---	---	---
	X20X36		---	---	---	---	---	---	---	*	---	---	*	*	---	*	---	---	---	*	---
	X23X35		---	---	---	---	---	---	---	---	---	---	---	---	---	*	---	---	---	---	---
	X28X32		---	---	---	---	---	---	---	*	---	---	*	*	---	---	---	---	---	---	---
	X43X44		*	*	---	*	*	*	---	---	---	---	---	---	---	---	---	---	---	---	---

Based on Table 5.19, it seems that the categories of variables that impact “annual source energy” (eQUEST output) are the siting options, the window system, and the ventilation system, and nly one of the variables from the foundation system. However, the siting options, the foundation system, and the window system affect eQUEST performance metrics only when MBF takes value of 20, and MOBN takes value of 5 or 10. In addition, MI does not take the value of 3, i.e., 3-factor interaction. The categories of variables that affect ATHENA outputs, are the foundation system, the wall system, the roof system, and the window system.

CHAPTER 6

Model Validation

In order to evaluate the treed regression models, the validation data set is used in this step. The PRESS value based on this testing dataset is, then, computed to determine if there is any significant difference between different situations regarding the type of variables in the tree and the limit to split the node for each response. The PRESS value is calculated manually in Microsoft Excel. The formula to calculate this number is as follows:

$$\text{PRESS value} = \frac{1}{n}\sum_{i=1}^{n} (y_i - \hat{y_i})^2$$

where n is the number of runs in the data set, y_i is the actual response for run number i based on the testing data set, and $\hat{y_i}$ is the predicted performance metrics related to ith run based on the testing data set. In addition, MARE is also considered in this study as another metric to compare the models. The calculation of MARE is as follows:

$$\text{MARE} = \frac{1}{n}\sum_{i=1}^{n} \frac{|y_i - \hat{y_i}|}{y_i}$$

where n is the number of runs in the data set, y_i is the actual response for run number i based on the testing data set, and $\hat{y_i}$ is the predicted performance metrics related to ith run based on the testing data set.

6.1 Treed Regression Models

The validation results based on PRESS and MARE values for Kung's design are shown in Table 6.1. As it can be seen in this table, for "annual source energy," by changing the limit to split the node in the tree from 67 to 68 cases, PRESS and MARE values decrease. This means that, although by considering a larger number of sample in each TN, the number of TNs does not change, PRESS and MARE values are indicating an improvement in the model. For "GWP," by considering only categorical variables to generate the tree, the

PRESS and MARE values are larger than the ones when the tree generation is based on all of the predictors.

Table 6-1 The PRESS value and MARE based on treed regression for Kung's design

Response	Variable type in tree generation	interaction	Limits to split the node	PRESS value	MARE
Annual resource energy	All	no	67	467.66	0.0565
		no	68	230.36	0.0378
GWP	All	no	51	24,588,853.66	0.0704
		no	52	4,957,361.27	0.0199
	Categorical	no	60	43,193,566.25	0.1078
		yes	60	43,193,566.25	0.1078
Non-renewable energy	All	no	64	13,758,272,325.83	0.1357
		no	65	18,150,757,029.00	0.1502
	Categorical	no	64	9,622,749,159.12	0.1027
		yes	64	9,596,684,342.01	0.1058
		no	65	8,921,325,160.63	0.1015
		yes	65	9,419,863,630.24	0.1058

In addition, it seems that for "GWP," by having the biggest tree (tree with 3 TNs) the model represents the smallest PRESS and MARE values, which indicates an improvement in the modeling. In addition, based on the results for the "non-renewable energy," it seems that when the tree generation is based on all of predictors, the better fitted model based on the lowest PRESS and MARE values is the tree with 5 TNs, which is the biggest tree between two cases. The only case that indicates that the tree with smaller number of TNs is potentially better is related to "non-renewable energy," when the tree generation is based on only the categorical variables. In addition, it seems that considering the interaction term in the regression part does not have any positive role in improving the models.

The PRESS and the MARE values for the cases based on Martinez's design are also shown in Table 6.2. These values were calculated using the testing dataset for Martinez's design, which was described in Section 4.3. Based on this table, for "annual source energy," when the tree generation is based on all of predictors, the lowest PRESS and MARE values indicate that the tree with a smaller minimum number of observations in each TN (=28) gives the better fitted model for this performance metrics, however, the number of TNs is 3 in both cases. In addition, this table shows that the model based on the tree generated considering the categorical variables only, although it only contains two TNs, performs better than the tree generated from all of predictors and 3 TNs. Thus, for "annual source energy," the lowest PRESS and MARE values belong to the fitted model when the tree generation is based on all of predictors with 3 TNs and the minimum number of observations in each TN is 28. The model generated based on only the categorical variables ranks second in the performance.

Table 6-2 The PRESS and MARE values based on the treed regression for Martinez's design

Response	Variables type in tree generation	interaction	Limits to split the node	PRESS value	MARE
Annual source energy	Categorical	No	60	412.0327973	0.0519
	All	No	57	194.8584875	0.0371
		No	58	546.0879412	0.0698
GWP	Categorical	No	60	977,154,681.18	0.3462
		Yes		10,101,932.37	0.0493
	All	No	69	18,280,648.99	0.0684
		No	70	15,125,312.63	0.0616
Non-renewable energy	Categorical	No	72	2,449,866,889.84	0.0517
		No	73	2,822,779,539.68	0.053
	All	No	72	4,667,654,299.51	0.0747
		No	73	6,139,525,620.04	0.0863

For “GWP” and based on the lowest PRESS and MARE values, it is seen that when the tree generation is based on all of predictors, there is an improvement when the size of the tree is smaller (with 3 TNs). In addition, only the categorical predictors are used in the tree generation, a large PRESS value indicates that the tree may not be generated from only categorical variables for “GWP,” unless the interaction term is considered. By considering the interaction term in regression modeling part, it seems that even the model is better than the model with the tree generation based on all of predictors for “GWP.”

Also, for “non-renewable energy,” when the tree generation is based on all predictors, the lowest PRESS and MARE values indicate that the model based on the bigger tree (with 5 TNs) has a better performance. However, the result in this table shows that for “non-renewable energy,” the model based on the tree generation with considering only the categorical variables performs better. Between the trees generated from categorical variables only, the model with 4 TNs, i.e., the bigger one, performs better.

6.2 MARS Models

The MARS method cannot operate on the data based on the categorical variables only, all the models that are fitted by this method on the data set are considering all the predictors. Based the lowest PRESS and MARE values (see table 6.3), it seems that by changing the maximum number of basis function, the degree of freedom of interactions, and the minimum number of observation between nodes, only ten models for “annual source energy” have been identified. Also, for “GWP,” only three models and for “non-renewable energy,” three models have been identified. In comparison with the models fitted by treed regression method (see Table 6.1), the PRESS values related to MARS method show the lowest numbers among “annual source energy” and “non-renewable energy.” This indicates that the models generated by MARS method are better fitted than the models fitted by treed regression for Kung's design for these two performance metrics. However, for "GWP," based on Kung's design, treed regression has the lowest PRESS value. This indicated that the models fitted by treed regression for "GWP," based on Kung's design are better fitted to the actual data, compared to the models fitted by MARS. Based on the MARS results, it seems that for the "GWP" and "non-renewable energy," which are calculated in ATHENA, with fewer basis functions, i.e., 20, and with no interactions in the model, and fewer number of observations between the nodes, i.e., 2, the better model is fitted on the data set.

Table 6-3 PRESS and MARE values for based on MARS for Kung's design

Response	MBF	MI	MOBN	MSE	R-square	Press value	MARE	Actual basis function
Annual source energy	20 mbf	3 MI	2 MIO	77.83285	0.90223	72.8335	0.0225	14
	50 mbf	3 MI	2 MIO	81.92143	0.89003	81.921	0.0235	13
	100 mbf	3 MI	2 MIO	81.92143	0.89003	81.921	0.0235	13
	50 mbf	2 MI	2 MIO	85.25916	0.8855	84.896	0.02317	24
	20 mbf	2 MI	2 MIO	86.23123	0.88424	86.2306	0.0231	16
	100 mbf	2 MI	2 MIO	87.27560	0.88284	87.276	0.0238	11
	50 mbf	2 MI	5 MIO	89.26897	0.88016	89.269	0.0246	34
	100 mbf	2 MI	5 MIO	90.06339	0.8791	90.063	0.0239	14
	20 mbf	2 MI	5 MIO	91.13701	0.85832	91.1356	0.0245	12
	20 mbf	3 MI	5 MIO	91.13701	0.87765	91.1356	0.0245	12
GWP	20 mbf	1 MI	2 MIO	21,083,381.790	0.68	21,083,403.960	0.076	5
	50 mbf	1 MI	2 MIO	23,194,891.556	0.65004	23,194,922.509	0.078245	4
	50 mbf	2 MI	10 MIO	30,702,919.794	0.53676	30,702,901.855	0.092353	5
Non-renewable energy	20 mbf	1 MI	2 MIO	4,670,169,749.300	0.707	4,670,164,583.890	0.0743	12
	50 mbf	1 MI	2 MIO	5,535,547,588.518	0.6531	5,535,536,499.248	0.08292	7
	100 mbf	3 MI	5 MIO	8,550,192,767.780	0.464	8,550,208,176.234	0.10831	2

According to the lowest PRESS and MARE values for the MARS models based on Martinez's design which is shown in Table 6.4, it seems that by changing the maximum number of basis functions, the degree of freedom of interactions, and the minimum number of observation between knots, only seven models for "annual source energy" have been identified. Also, for "GWP," only six models and for "non-renewable energy," six models have been identified. Compared to the models fitted by treed regression method (see Table 6.2), the PRESS values calculated from the MARS method show the lowest numbers among all of three responses. This means that the models fitted by MARS have a better fit compared to the models fitted by the treed regression based on Martinez's design.

Based on the MARS results in Table 6.4, it seems that for eQUEST output, with more minimum observations between knots. i.e., 10, and by considering interactions in the model; MI takes values of 2 or 3, and when maximum number of basis functions takes values of 20 and 100, a better model is fitted on the dataset. For "GWP" and "non-renewable energy," which are from ATHENA, by considering two-factor interactions or even no interaction in modeling, the fitted model performs better. For "non-renewable energy" with fewer number of minimum observations between knots; i.e., value of 2, and fewer number for the maximum basis functions, a better model is fitted on the data set. This is also true for "GWP," however, the model fitted by considering the minimum observations between knots as 10, performs best.

Table 6-4 PRESS and MARE values for based on MARS for Martinez's design

Response	MBF	MI	MOBN	MSE	R-square	Press value	MARE	Actual Basis Function
Annual source energy	20 mbf	2 MI	10 MIO	167.207	0.625	167.203	0.033	13
	100 mbf	3 MI	10 MIO	183.383	0.588	183.382	0.033	2
	20 mbf	1 MI	10 MIO	210.098	0.528	210.100	0.034	12
	100 mbf	2 MI	10 MIO	227.440	0.489	227.441	0.034	2
	20 mbf	2 MI	2 MIO	228.474	0.487	228.474	0.036	2
	20 mbf	2 MI	5 MIO	232.439	0.478	232.435	0.036	2
	20 mbf	1 MI	5 MIO	248.932	0.441	248.937	0.037	10
GWP	20 mbf	2 MI	10 MIO	6,235,068.560	0.891	6,235,021.582	0.039	11
	100 mbf	1 MI	2 MIO	6,273,706.326	0.890	6,273,688.075	0.041	18
	50 mbf	1 MI	2 MIO	6,496,416.493	0.886	6,496,437.764	0.041	17
	20 mbf	3 MI	2 MIO	6,740,348.552	0.882	6,740,310.980	0.041	10
	50 mbf	2 MI	10 MIO	6,767,850.533	0.881	6,767,863.221	0.041	10
	20 mbf	1 MI	2 MIO	6,991,231.009	0.877	6,991,264.269	0.042	10
Non-renewable energy	20 mbf	2 MI	2 MIO	1,254,254,161.542	0.914	1,254,257,148.388	0.039	14
	20 mbf	1 MI	2 MIO	1,289,652,813.432	0.912	1,289,652,089.441	0.038	11
	100 mbf	1 MI	2 MIO	1,330,134,368.675	0.909	1,330,133,786.565	0.039	13
	50 mbf	1 MI	2 MIO	1,337,946,583.749	0.908	1,337,946,461.000	0.042	26
	100 mbf	3 MI	5 MIO	1,352,204,539.671	0.907	1,352,210,770.304	0.040	10
	50 mbf	3 MI	5 MIO	1,370,311,066.165	0.907	1,370,319,482.593	0.039	8

6.3 Discussion

Based on the results in Tables 6.1 and 6.2, that the first performance metric "Annual source energy" has the lowest PRESS value with the treed regression model based on Martinez's design when all of the predictors are considered in the tree generation part, using 57 cases as the limitation to split the nodes. The second performance metric "GWP" has the lowest PRESS value with the treed regression approach based on Kung's design when all of the predictors are considered in the tree generation part, using 52 cases as the limitation to split the nodes. Finally, the third performance metric "Non-renewable energy" has the lowest PRESS value with the treed regression approach based on Martinez's design when only categorical predictors are considered in the tree generation part, using 72 cases as the limitation to split the nodes.

Based on the results of the PRESS value for the MARS approach for each design in Tables 6.3 and 6.4, the first performance metric "Annual source energy" has the lowest PRESS value with the MARS approach based on Kung's design when MBF is 20, MI is 3, and MOBN is 2. The second performance metric

"GWP" has the lowest PRESS value with the MARS approach based on Martinez's design when MBF is 20, MI is 2, and MOBN is 10. Finally, the third performance metric "Non-renewable energy" has the lowest PRESS value with the MARS approach based on Kung's design when MBF is 20, MI is 2, and MOBN is 2.

If only one model is selected as the best model for each performance metric among all of the cases, the following statements will be the best summary. For the first performance metric "Annual source energy" from eQUEST, the best model based on the lowest PRESS value is based on the MARS approach with Kung's design. The best model for the second performance metric "GWP" from ATHENA is achieved when the treed regression approach and Kung's design are used. Finally, for the third performance metric

“Non-renewable energy” from ATHENA, the best model is achieved when the MARS approach and Martinez’s design are used.

Further, by considering the two designs used in this study based on the MARS approach, some interesting points can be concluded. As it was mentioned in previous sections, Kung’s design handles the categorical variables using MA and continuous variables using a Sobol’ sequence and combines them in a single design using a Latin hypercube design. This approach provides a strong advantage for Kung’s design, since it treats different type of variables separately, and a disadvantage for Kung’s design, since it may not represent the interactions between the categorical and the continuous variables.

Martinez’s design, on the other hand, considers all of the variables as continuous. This property provides an advantage for this approach since it is balanced over the entire space by using the Sobol’ sequence, and as a shortcoming, since it does not treat two different types of variables using separate methods, which may be more appropriate for that specific type of variable. Based on these differences between Kung’s design and Martinez’s design, it is expected that Martinez’s design performs better in modelling the interactions between the categorical and the continuous variables, if they exist. Since the continuous variables are only available in eQUEST, “annual source energy” is the only performance metrics that may include interactions between two types of variables in the respective fitted models. Based on Table 5.18, the models for “annual source energy” based on the MARS approach and Kung’s design include “X32X43,”,” X38X42, X20X41, X15X41, X10X43, X6X40, which represent the interactions between the categorical and the continuous variables. However, the best model for “annual source energy” based on the MARS approach and Kung’s design, which has MBF=20, MI=3, and MOBN=2 (see Table 6.3), does not include any interaction terms between the categorical and the continuous variables. In addition, based on Table 5.19, the models for “annual source energy" based on the MARS approach and Martinez's design do not include any interaction terms between the categorical and the continuous variables, except the best model with MBF=20, MI=2, and MOBN=10 (see Table 6.4). Based on the results in Table 5.19, this model includes the interaction between "Foot print shape-X1" and "Ventilation – bedroom-X40," which represents an interaction between the categorical and the continuous variables. Therefore, it seems that based on the MARS approach, Kung's design performs better in modeling the interaction between the categorical and the continuous variables compared to Martinez's design, which was not expected.

CHAPTER 7

CONCLUSION AND FUTURE WORK

As a conclusion, it seems that either of the two designs can handle a part of the several situations considered in this study, for either of the performance metrics from eQUEST and ATHENA. For example, using Martinez's design as an experimental design framework for decision variables, better models are fitted on the dataset of "non-renewable energy," regardless of the implemented statistical approach. However, for "GWP," either of the two designs can work well. When the treed regression is considered for the statistical analysis, Martinez's design performs better, and when the MARS method is used, the performance of Kung's design is better. In addition, Kung's design performs better for "annual source energy," when the MARS approach is used as the statistical analysis, while Martinez's design performs better when the treed regression is used.

Because of the limited duration of this study, only 192 runs were considered for the training dataset and 96 runs for the testing data set, which may be inadequate for the complexity of green building or other real world applications. Thus, for future work, it is recommended to consider a larger number of runs, which may provide more accurate results. Also, among of several options for variable levels in eQUEST or ATHENA, only some of them were selected in this study. It is recommended to select other options or select more levels for predictor variables, to investigate if any important level is neglected.

With regard to future work in design of experiments, Kung's design is a novel hybrid approach that has shown promise for this complex green building case study. As was mentioned in Section 6.3, this study shows an unexpected result for the modeling of the interactions between the categorical and the continuous variables for two designs, where Pin's design performed better than Martinez's design. Therefore, there are two design-related issues for future work. First, create a dataset to better study the impact on interactions between categorical and continuous variables. Second, explore alternate methods to the Latin hypercube design for combining the MA for discrete variables and the Sobol' sequence for continuous variables.

Finally, with regard to statistical modeling, it is recommended to use a multiple-response modeling approach, for example, the seemingly unrelated regressions (SUR) method [37]. Since the performance metrics in this study are correlated, the SUR method can be used to improve the precision of the model. Further, by fitting a multiple-response model, it is possible to incorporate all responses into a single model.

Appendix A

Kung’s design for 96-point testing dataset

Table 0.1 Kung's design for 96-point testing dataset

run #	OA	S	run #	OA	S	run #	OA	S
1	13	49	33	87	51	65	86	41
2	35	55	34	65	12	66	68	87
3	54	94	35	83	8	67	49	2
4	11	27	36	89	6	68	82	56
5	72	69	37	22	71	69	90	90
6	51	48	38	17	36	70	93	46
7	39	47	39	20	75	71	45	60
8	76	35	40	50	64	72	43	22
9	23	39	41	3	43	73	84	66
10	16	54	42	52	52	74	77	61
11	46	59	43	48	79	75	32	7
12	61	58	44	53	73	76	63	93
13	34	83	45	12	42	77	62	68
14	81	16	46	67	25	78	95	91
15	44	38	47	21	30	79	33	24
16	10	4	48	15	57	80	36	29
17	96	26	49	14	70	81	91	85
18	1	14	50	71	13	82	25	86
19	92	76	51	47	78	83	18	77
20	2	3	52	38	20	84	24	1
21	9	81	53	4	19	85	74	23
22	73	82	54	26	63	86	94	45
23	30	9	55	19	15	87	31	72
24	59	28	56	28	18	88	37	21
25	7	44	57	55	62	89	57	50
26	75	67	58	27	5	90	42	32
27	60	17	59	29	80	91	69	11
28	56	40	60	8	65	92	80	37
29	85	10	61	64	53	93	88	74
30	6	92	62	66	84	94	5	89
31	40	95	63	78	34	95	58	31
32	70	96	64	79	33	96	41	88

Appendix B

MARS results based on Kung's and Martinez's designs

Table 0.2 MARS results based on Kung's design

R	MBF	MI	MOBN	MSE	R-square	Press value	MARE	ABF
1	20	1	2	156.31	0.790	156.3128	0.027	13
1	20	1	5	152.37	0.795	152.3693	0.030	14
1	20	1	10	147.92	0.801	147.9215	0.027	12
1	20	2	2	86.23	0.884	86.2306	0.023	16
1	20	2	5	91.14	0.858	91.1356	0.025	12
1	20	2	10	105.54	0.878	105.5391	0.026	12
1	20	3	2	77.83	0.902	72.8335	0.023	14
1	20	3	5	91.14	0.878	91.1356	0.025	12
1	20	3	10	121.36	0.837	129.3630	0.027	16
1	50	1	2	141.70	0.810	12501.6310	0.315	38
1	50	1	5	146.37	0.804	146.3670	0.028	10
1	50	1	10	138.34	0.814	599.6300	0.071	25
1	50	2	2	85.26	0.886	84.8960	0.023	24
1	50	2	5	89.27	0.880	89.2690	0.025	34
1	50	2	10	100.29	0.865	110.3050	0.026	13
1	50	3	2	81.92	0.890	81.9210	0.024	13
1	50	3	5	91.25	0.878	108.3280	0.028	12
1	50	3	10	109.60	0.853	755.1720	0.059	18
1	100	1	2	139.31	0.813	139.3040	0.028	13
1	100	1	5	132.74	0.822	132.7380	0.027	38
1	100	1	10	144.28	0.806	144.2780	0.026	23
1	100	2	2	87.28	0.883	87.2760	0.024	11
1	100	2	5	90.06	0.879	90.0630	0.024	14
1	100	2	10	100.29	0.865	100.2880	0.025	13
1	100	3	2	81.92	0.890	81.9210	0.024	13
1	100	3	5	94.79	0.873	846.5710	0.058	26
1	100	3	10	106.36	0.857	106.3545	0.026	10
2	20	1	2	21083381.79	0.680	21083403.9600	0.076	5
2	20	1	5	21083381.79	0.680	21083403.9600	0.076	5
2	20	1	10	21083381.79	0.680	21083403.9600	0.076	5
2	20	2	2	32352915.33	0.511	32352972.1600	0.096	4
2	20	2	5	32352915.33	0.511	32352972.1600	0.096	4
2	20	2	10	32352915.33	0.511	32352972.1600	0.096	4
2	20	3	2	36529050.66	0.448	49410995.7400	0.109	2
2	20	3	5	36529050.66	0.448	49410995.7400	0.109	2

2	20	3	10	36529050.66	0.448	49410995.7400	0.109	2
2	50	1	2	23194891.56	0.650	23194922.5085	0.078	4
2	50	1	5	23194891.55	0.650	23194922.5085	0.078	4
2	50	1	10	23194891.56	0.650	23194922.5085	0.078	4
2	50	2	2	3380255.18	0.490	33802551.3796	0.096	1
2	50	2	5	3380255.18	0.490	33802551.3796	0.096	1
2	50	2	10	30702919.79	0.537	30702901.8547	0.092	5
2	50	3	2	36529050.00	0.448	36529121.5985	0.096	2
2	50	3	5	36529050.00	0.448	36529121.5985	0.096	2
2	50	3	10	36529050.00	0.448	36529121.5985	0.096	2
2	100	1	2	23194891.56	0.650	23194922.5085	0.078	4
2	100	1	5	23194891.55	0.650	23194922.5085	0.078	4
2	100	1	10	23194891.56	0.650	23194922.5085	0.078	4
2	100	2	2	3380255.18	0.490	33802551.3796	0.096	1
2	100	2	5	3380255.18	0.490	33802551.3796	0.096	1
2	100	2	10	3380255.18	0.490	33802551.3796	0.096	1
2	100	3	2	3380255.18	0.490	36529121.5985	0.096	2
2	100	3	5	36529050.00	0.448	36529121.5985	0.096	2
2	100	3	10	3380255.18	0.490	36529121.5985	0.096	2
3	20	1	2	4670169749.30	0.707	4670164583.8900	0.074	12
3	20	1	5	4670169749.30	0.707	4670164583.8900	0.074	12
3	20	1	10	4670169749.30	0.707	4670164583.8900	0.074	12
3	20	2	2	9637766614.56	0.396	21997253553.2324	0.161	1
3	20	2	5	9637766614.56	0.396	21997253553.2324	0.161	1
3	20	2	10	9115875863.52	0.428	9115865327.2810	0.110	3
3	20	3	2	9637766614.56	0.396	21997253553.2324	0.161	1
3	20	3	5	9637766614.56	0.396	21997253553.2324	0.161	1
3	20	3	10	9637766614.56	0.396	21997253553.2324	0.161	1
3	50	1	2	5535547588.52	0.653	5535536499.2480	0.083	7
3	50	1	5	5535547588.52	0.653	5535536499.2480	0.083	7
3	50	1	10	5535547588.52	0.653	5535536499.2480	0.083	7
3	50	2	2	9637766614.56	0.396	9637773667.4852	0.114	1
3	50	2	5	9637766614.56	0.396	9637773667.4852	0.114	1
3	50	2	10	11179454516.49	0.299	11179458381.4810	0.119	6
3	50	3	2	9637766614.56	0.396	9637773667.4852	0.114	1
3	50	3	5	9637766614.56	0.396	9637773667.4852	0.114	1
3	50	3	10	9637766614.56	0.396	9637773667.4852	0.114	1
3	100	1	2	5535547588.52	0.653	5535536499.2480	0.083	7
3	100	1	5	5535547588.52	0.653	5535536499.2480	0.083	7

3	100	1	10	5535547588.52	0.653	5535536499.2480	0.083	7
3	100	2	2	9637766614.56	0.396	9637773667.4852	0.114	1
3	100	2	5	8550192767.78	0.464	9637773667.4852	0.114	1
3	100	2	10	10637927427.46	0.333	10637921408.7600	0.114	5
3	100	3	2	9637766614.56	0.396	9637773667.4852	0.114	1
3	100	3	5	8550192767.78	0.464	8550208176.2340	0.108	2
3	100	3	10	9637766614.56	0.396	9637773667.4852	0.114	1

Table 0.3 MARS results based on Martinez's design

R	MBF	MI	MOBN	MSE	R-square	Press value	MARE	ABF
1	20	1	2	272.59	0.388	0.0396	272.590	9
1	20	1	5	248.93	0.441	0.0370	248.937	10
1	20	1	10	210.10	0.528	0.0343	210.100	12
1	20	2	2	228.47	0.487	0.0357	228.474	2
1	20	2	5	232.44	0.478	0.0358	232.435	2
1	20	2	10	167.21	0.625	0.0335	167.203	13
1	20	3	2	228.47	0.487	0.0357	228.474	2
1	20	3	5	232.44	0.478	0.0358	232.435	2
1	20	3	10	167.21	0.625	0.0335	167.203	13
1	50	1	2	296.34	0.334	0.0445	296.143	4
1	50	1	5	276.80	0.378	0.0415	276.799	9
1	50	1	10	275.02	0.382	0.0396	275.022	3
1	50	2	2	228.47	0.486	0.0357	228.474	2
1	50	2	5	232.43	0.478	0.0358	232.446	2
1	50	2	10	220.85	0.504	0.0424	270.178	7
1	50	3	2	228.47	0.480	0.0357	228.474	2
1	50	3	5	232.43	0.478	0.0358	232.446	2
1	50	3	10	260.19	0.415	0.0404	260.203	6
1	100	1	2	296.35	0.335	0.0445	296.350	4
1	100	1	5	279.93	0.371	0.0394	279.932	3
1	100	1	10	282.37	0.366	0.0434	282.366	7
1	100	2	2	228.47	0.487	0.0357	228.474	2
1	100	2	5	232.44	0.478	0.0358	232.435	2
1	100	2	10	227.44	0.489	0.0340	227.441	2
1	100	3	2	228.47	0.487	0.0357	228.474	2
1	100	3	5	232.44	0.478	0.0358	232.435	2
1	100	3	10	183.38	0.588	0.0335	183.382	2
2	20	1	2	6991231.01	0.877	0.0416	6991264.269	10
2	20	1	5	6991231.01	0.877	0.0416	6991264.269	10
2	20	1	10	6991231.01	0.877	0.0416	6991264.269	10
2	20	2	2	7275539.27	0.872	0.0436	7275560.333	10
2	20	2	5	7275539.27	0.872	0.0436	7275560.333	10
2	20	2	10	6235068.56	0.891	0.0393	6235021.582	11
2	20	3	2	6740348.55	0.882	0.0414	6740310.980	10
2	20	3	5	7275539.27	0.872	0.0436	7275560.333	10
2	20	3	10	6235068.56	0.891	0.0393	6235021.582	10

2	50	1	2	6496416.49	0.886	0.0415	6496437.764	17
2	50	1	5	6496416.49	0.886	0.0415	6496437.764	17
2	50	1	10	6496416.49	0.886	0.0415	6496437.764	17
2	50	2	2	8485168.65	0.851	0.0481	8485157.090	9
2	50	2	5	8485168.65	0.851	0.0481	8485157.090	9
2	50	2	10	6767850.53	0.881	0.0411	6767863.221	10
2	50	3	2	8659448.20	0.848	0.0481	8659358.691	10
2	50	3	5	7593111.54	0.867	0.0448	7593099.469	16
2	50	3	10	8968119.49	0.843	0.0471	8968087.473	23
2	100	1	2	6273706.33	0.890	0.0410	6273688.075	18
2	100	1	5	6273706.33	0.890	0.0410	6273688.075	18
2	100	1	10	6273706.33	0.890	0.0410	6273688.075	18
2	100	2	2	8023324.49	0.859	0.0462	8023312.202	16
2	100	2	5	8023324.49	0.859	0.0462	8023312.202	16
2	100	2	10	7674067.96	0.865	0.0455	7674095.854	15
2	100	3	2	7174837.94	0.874	0.0442	7174843.573	11
2	100	3	5	9280990.35	0.837	0.0489	9280998.151	11
2	100	3	10	7672211.15	0.865	0.0437	7672219.404	19
3	20	1	2	1289652813.43	0.912	0.0377	1289652089.441	11
3	20	1	5	1289652813.43	0.912	0.0377	1289652089.441	11
3	20	1	10	1289652813.43	0.912	0.0377	1289652089.441	11
3	20	2	2	1254254161.54	0.914	0.0391	1254257148.388	14
3	20	2	5	1254254161.54	0.914	0.0391	1254257148.388	14
3	20	2	10	1254254161.54	0.914	0.0391	1254257148.388	14
3	20	3	2	1387037222.34	0.905	0.0409	1387036628.107	12
3	20	3	5	1387037222.34	0.905	0.0409	1387036628.107	12
3	20	3	10	1387037222.34	0.905	0.0409	1387036628.107	12
3	50	1	2	1337946583.75	0.908	0.0417	1337946461.000	26
3	50	1	5	1337946583.75	0.908	0.0417	1337946461.000	26
3	50	1	10	1337946583.75	0.908	0.0417	1337946461.000	26
3	50	2	2	1713079161.55	0.883	0.0456	1713079598.000	17
3	50	2	5	1713079161.55	0.883	0.0456	1713079598.000	17
3	50	2	10	1713079161.55	0.883	0.0456	1713079598.000	17
3	50	3	2	1715868223.25	0.883	0.0427	1715862872.858	9
3	50	3	5	1370311066.16	0.907	0.0387	1370319482.593	8
3	50	3	10	1536556595.68	0.895	0.0418	1536563338.993	9
3	100	1	2	1330134368.67	0.909	0.0391	1330133786.565	13
3	100	1	5	1330134368.67	0.909	0.0391	1330133786.565	13
3	100	1	10	1330134368.67	0.909	0.0391	1330133786.565	13

3	100	2	2	1908812783.97	0.869	0.0470	1908820415.000	13
3	100	2	5	1908812783.97	0.869	0.0470	1908820415.000	13
3	100	2	10	1813897130.45	0.876	0.0462	1813898602.000	13
3	100	3	2	1686183788.37	0.885	0.0446	1686174353.000	17
3	100	3	5	1352204539.67	0.907	0.0398	1352210770.304	10
3	100	3	10	1429905024.98	0.903	0.0401	1429905622.602	9

Appendix C

Fitted Models

Table 0.4-Treed regression model for “annual source energy,”,” by considering all variables in tree generation and at least 67 cases to split TN, based on Kung’s design

TN #	Fitted Model
1	[332.95+110.38566×X_{38} -5.49438×X_{43} +3.12957×X_{44}].|{X_{43}<=22.06}
2	[251.78626+50.93673×X_{36}+51.98230×X_{37}+57.85489×X_{38}+0.68133×X_{44}].|{X_{43}>22.06}.|{X_1:b}].|{X_{43}<=64.52}
3	[281.886].|{X_{43}>22.06}.|{X_1:b}.|{X_{43}>64.52}
4	[300.11746-0.68905×X_{10}-0.37125×X_{21}+80.93224×X_{36}+44.50631×X_{37}-0.09863×X_{43}].|{X_{43}>22.06}.|{X1:a}.|{X_2:a,b}
5	[299.65527-0.72303×X_6-1.82532×X_9+172.30727×X_{38}].|{X_{43}>22.06}.|{X1:a}.|{X_2:c,d}

Table 0.5-Treed regression model for “annual source energy,”,” by considering all variables in tree generation and at least 68 cases to split TN, based on Kung’s design

TN #	Fitted Model
1	[333.81540+118.99405×X_{38}-4.57465×X_{43}+2.55243×X_{44}].|{X_{43}<=24.44}
2	[251.70796+50.74322×X_{36}+52.29327×X_{37}+57.58624×X_{38}+0.68553×X_{44}].|{X_{43}>24.44}.|{X_1:b}] .|{X_{43}<=64.52}
3	[281.886].|{X_{43}>24.44}.|{X_1:b}.|{X_{43}>64.52}
4	[295.58275-0.72122×X_{10}-0.45057×X_{21}+78.27395×X_{36}-44.54714×X_{37}].|{X_{43}>24.44}.|{X1:a}.|{X_2:a,b}
5	[298.67769-0.67541×X_6 -1.71943×X_9+167.69132×X_{38}].|{X_{43}>24.44}.|{X1:a}.|{X_2:c,d}

Table 0.6-Treed regression model for “GWP” without interaction, by considering only the categorical variables in tree generation and at least 60 cases to split TN, based on Kung’s design

TN #	Fitted Model
1	[2490.40044+655.53021×X_3+2650.38159×X_4+1.50988×X_5+99.55576×X_{19}+18153×X_{27}+19377×X_{35} +16692×X_{37}].|{X_{23}:c}
2	[17692 + 1287.15473×X_3+2240.47942×X_4+356.47504×X_6+423.93251×X_{20}].|{X_{23}:a,b}

Table 0.7-Treed regression model for “GWP” with interaction, by considering only the categorical variables in tree generation and at least 60 cases to split TN, based on Kung’s design

TN #	Fitted Model
1	[2490.40044+655.53021×X_3+2650.38159×X_4+1.50988×X_5+99.55576×X_{19}+18153×X_{27}+19377×X_{35} +16692×X_{37}].|{X_{23}:c}
2	[17692 + 1287.15473×X_3+2240.47942×X_4+356.47504×X_6+423.93251×X_{20}].|{X_{23}:a,b}

Table 0.8-Treed regression model for “GWP,”,” by considering all variables in tree generation and at least 51 cases to split TN, based on Kung’s design

TN #	Fitted Model
1	[27452 + 349.79227×X_6+380.25655×X_{20}+52280×X_{35}].|{X_4<=6}
2	[29679 + 3.26075×X_5+131.24888×X_6+85.45087×X_{19}+188.96234×X_{20}+8719.03403×X_{27}].|{ X_4>6}

Table 0.9-Treed regression model for”GWP,”,” by considering all variables in tree generation and at least 52 cases to split TN, based on Kung’s design

TN #	Fitted Model
1	$[27452 + 349.79227\times X_6+380.25655\times X_{20}+52280\times X_{35}].\|\{X_4<=6\}$
2	$[30910 + 3.08573\times X_5+484.95781\times X_8+90.09806\times X_{19}+36482\times X_{35}+18335\times X_{38}].\|\{X_4>6\}.\|\{X_5<=4500\}$
3	$[36380 + 2.46679\times X_5+204.10905\times X_6+252.60902\times X_{20}+15721\times X_{36}].\|\{X_4>6\}.\|\{X_5>4500\}$

Table 0.10-Treed regression model for "non-renewable energy,"," by considering all variables in tree generation and at least 64 cases to split TN, based on Kung's design

TN #	Fitted Model
1	$[582207 + 6244.16331\times X_3+2959.22066\times X_6+6409.35355\times X_{10}+3382.28574\times X_{20}-208708\times X_{35}].\|\{X_{23}:a,b\}.\|\{X_6<=24.5\}.\|\{X_4<=6\}$
2	$[502809+37.70944\times X_5+379591\times X_{35}].\|\{X_{23}:a,b\}.\|\{X_6<=24.5\}.\|\{X_4>6\}$
3	$[776391.533].\|\{X_{23}:a,b\}.\|\{X_6>24.5\}$
4	$[499098+7457.86579\times X_{10}+450519\times X_{27}].\|\{X_{23}:c\}.\|\{X_4<=6\}$
5	$[695840+1743.17484\times X_{19}+1325.71275\times X_{20}+235468\times X_{27}+136814\times X_{37}].\|\{X_{23}:c\}.\|\{X_4>6\}$

Table 0.11-Treed regression model for "non-renewable energy,"," by considering all variables in tree generation and at least 65 cases to split TN, based on Kung's design

TN #	Fitted Model
1	$[582207 + 6244.16331\times X_3+2959.22066\times X_6+6409.35355\times X_{10}+3382.28574\times X_{20}-208708\times X_{35}].\|\{X_{23}:a,b\}.\|\{X_6<=24.5\}.\|\{X_4<=6\}$
2	$[502809+37.70944\times X_5+379591\times X_{35}].\|\{X_{23}:a,b\}.\|\{X_6<=24.5\}.\|\{X_4>6\}$
3	$[776391.533].\|\{X_{23}:a,b\}.\|\{X_6>24.5\}$
4	$[506581+20585\times X_4+2157.23644\times X_{19}+329503\times X_{27}].\|\{X_{23}:c\}$

Table 0.12-Treed regression model for "non-renewable energy" without interactions, by considering only the categorical variables in tree generation and at least 64 cases to split TN, based on Kung's design

TN #	Fitted Model
1	$[354594 + 31555\times X_4+2203.70903\times X_6+2148.89651\times X_{20}].\|\{X_{23}:a,b\}.\|\{X_{14}:a\}.\|\{X_{23}:a\}$
2	$[161047+18318\times X_4+17.21478\times X_5+12753\times X_9+302.04139\times X_{18}+2713.95393\times X_{19}+1795.94426\times X_{20}+172804\times X_{27}-14929\times X31+145420\times X_{35}+589614\times X_{36}].\|\{X_{23}:a,b\}.\|\{X_{14}:a\}.\|\{X_{23}:b\}$
3	$[596083+8017.91426\times X_{20}].\|\{X_{23}:a,b\}.\|\{X_{14}:b\}$
4	$[539132+26203\times X_4+17357\times X_8+1645.16783\times X_{10}+918.53809\times X_{19}+267416\times X_{36}+224826\times X_{37}].\|\{X_{23}:c\}.\|\{X_1:b\}$
5	$[777170 + 1751.07406\times X_{18}].\|\{X_{23}:c\}.\|\{X_1:a\}$

Table 0.13-Treed regression model for "non-renewable energy" with interactions, by considering only the categorical variables in tree generation and at least 64 cases to split TN, based on Kung's design

TN #	Fitted Model
1	$[354594 + 31555\times X_4+2203.70903\times X_6+2148.89651\times X_{20}].\|\{X_{23}:a,b\}.\|\{X_{14}:a\}.\|\{X_{23}:a\}$
2	$[161047+18318\times X_4+17.21478\times X_5+12753\times X_9+302.04139\times X_{18}+2713.95393\times X_{19}+1795.94426\times X_{20}+172804\times X_{27}-14929\times X31+145420\times X_{35}+589614\times X_{36}].\|\{X_{23}:a,b\}.\|\{X_{14}:a\}.\|\{X_{23}:b\}$

3	[596083+8017.91426× X_{20}].|{ X_{23}:a,b}.|{ X_{14}:b}
4	[575079+26744× X_4-13562×std X_4X_6+15115× X_8+1644.91291× X_{10}+342772× X_{36}].|{ X_{23}:c}.|{ X_1:b}
5	[777170 + 1751.07406× X_{18}].|{ X_{23}:c}.|{ X_1:a}

Table 0.14-Treed regression model for "non-renewable energy" without interactions, by considering only the categorical variables in tree generation and at least 65 cases to split TN, based on Kung's design

TN #	Fitted Model
1	[199965 + 16022× X_3 + 20924× X_4 +21.92953× X_5 +1451.62142× X_6 +4205.69180× X_9 +6375.71436× X_{10} +2028.63769× X_{19}] .|{ X_{23}:a,b}.|{ X_{14}:a}
2	[596083+8017.91426× X_{20}].|{ X_{23}:a,b}.|{ X_{14}:b}
3	[506581 + 20585× X_4+2157.23644× X_{19}+329503× X_{27}].|{ X_{23}:c}

Table 0.15-Treed regression model for "non-renewable energy" with interactions, by considering only the categorical variables in tree generation and at least 65 cases to split TN, based on Kung's design

TN #	Fitted Model
1	[285635 + 14388× X_3 +18010× X_4 +18.78362× X_5 -21498×std X_4X_6 +1888.09063× X_6 + 4125.20321× X_{10} +1952.19729× X_{19} - 81997× X_{27} +154455× X_{35} +193105× X_{38}].|{ X_{23}:a,b}.|{ X_{14}:a}
2	[596083+8017.91426× X_{20}].|{ X_{23}:a,b}.|{ X_{14}:b}
3	[506581 + 20585× X_4+2157.23644× X_{19}+329503× X_{27}].|{ X_{23}:c}

Table 0.16-MARS model for "annual source energy,"," with MBF=20, MI=3, and MOBN=2, based on Kung's design

188.383 - 8.25167 *MAX (0, X43-22.4603) + 9.79774 *MAX(0,22.4603-X43) + 0.237827 *MAX (0, X44-10)*MAX(0, X43-22.4603)+ 9.4845 *(.|{X1 : a}) + 44.4801 *MAX (0, X38-0.06) - 0.388443 *MAX (0, X6-9.53674E-007).|{X1 : a}+ 8.35027 .|{X2 : c, d} .|{X1 : a}- 1.0996 *MAX (0, X10-0) + 8.0891 *MAX(X43-9.7619)-(0.663387 *MAX(0,25.2381-X44)+294.283 *MAX (0, X38-0.18).|{X2 : c, d} + 5.78999 *MAX(0,X35-0.06)*MAX(0, X10-0)+ 0.701153 *MAX (0, 11-X6)+ 28.8696 *MAX (0, X36-0.06)

Table 0.17-MARS model for "annual source energy,"," with MBF=50, MI=3, and MOBN=2, based on Kung's design

185.477 - 9.18638 *max (0, X43 - 22.4603) + 10.379 *max (0, 22.4603 - X43) + 0.235295 *max (0, X44 - 10)* max (0, 22.4603 - X43)+ 11.5497 .|{X1 : a} - 0.453165 * max (0, X6 - 0).|{X1 : a}+ 8.48317 .|{X2 : c, d} .|{X1 : a}- 1.0966 * max(0, X10 - 0) + 8.9842 * max(0,X43 - 9.7619) - 0.646235 * max(0, 25.2381 -X44)+ 378.715 * max(0, X38 - 0.18) .|{X2 : c, d}) .|{X1 : a}+ 5.5073 * max(0, X35 - 0.06) * max(0, X10 - 0) + 0.666917 * max(0, 11 -X6)- 1.9676 .|{X33 : d} * max(0, X9 - 4) .|{X1 : a}

Table 0.18-MARS model for "annual source energy,"," with MBF=100, MI=3, and MOBN=2, based on Kung's design

185.477 - 9.18638 * max(0, X43 - 22.4603) + 10.379 * max(0, 22.4603 -X43) + 0.235295 * max(0, X44 - 10) * max(0, 22.4603 -X43)+ 11.5497 .|{X1 : a}- 0.453165 * max(0,X6 - 0) *.|{X1 : a} + 8.48317 .|{X2 : c, d}.|{X1 : a}- 1.0966 * max(0, X10 - 0) + 8.9842 * max(0, X43 - 9.7619) - 0.646235 * max(0, 25.2381 -X44)+ 378.715 * max(0, X38 - 0.18) .|{X2 : c, d} .|{X1

: a}+ 5.5073 * max(0, X10 - 0)+ 0.666917 * max(0, 11 -X6)- 1.9676 .|{X33 : d} * max(0, X9 - 4) .|{X1 : a}

Table 0.19-MARS model for “annual source energy,”,” with MBF=50, MI=2, and MOBN=2, based on Kung’s design

81.645 - 8.94169 * max(0, X43 - 22.4603)+ 9.14848 * max(0, 22.4603 -X43)+ 0.277654 * max(0, X44 - 10) *max(0, 22.4603 -X43)+ 8.79774 .|{X1 : a}+ 67.2371 * max(0, X38 - 0.06)- 0.432442 * max(0,X6 - 0) .|{X1 : a}+ 11.0903 .|{X2 : c, d} .|{X1 : a}- 0.527013 * max(0, X10 - 0)+ 8.6792 * max(0, X43 - 9.7619)- 1.21661 * max(0, 25.2381 -X44)+ 1.34908 * max(0, 11 - X6)- 0.17074 * max(0, X3 - 4) * max(0, 11 -X6)- 0.333293 .|{X24 : a} * max(0, X10 - 0)- 702.361 * max(0, X36 - 0.18) * max(0, X38 - 0.06)- 3.69258 * max(0, X9 - 4) * max(0, X38 - 0.06)+ 0.0169412 * max(0, X43 - 38.3333) * max(0, 25.2381 -X44)+ 0.0437787 * max(0, 38.3333 -X43) * max(0, 25.2381 -X44)- 0.483397 .|{X33 : d} * max(0, X10 - 0)+ 5.47189 .|{X15 : b} .|{X1 : a}+ 3.41049 * max(0, X35 - 0.06) * max(0, X10 - 0)+ 45.8801 * max(0, X36 - 0.06)+ 0.472049 * max(0,X37 - 0.06) * max(0, X43 - 22.4603)- 0.0279807 * max(0, X7 - 2) * max(0, X43 - 9.7619)+ 102.567 .|{X2 : c, d} * max(0, X38 - 0.18)

Table 0.20-MARS model for “annual source energy,”,” with MBF=20, MI=2, and MOBN=2, based on kung’s design

189.092 - 8.51582 * max(0, X43 - 22.4603)+ 9.84827 * max(0, 22.4603 -X43)+ 0.236745 * max(0, X44 - 10) * max(0, 22.4603 -X43)+ 10.7972 .|{X1 : a}+ 67.4912 * max(0, X38 - 0.06)- 0.436428 * max(0,X6 -0) .|{X1 : a}+ 12.9183 .|{X2 : c, d}.|{X1 : a}- 0.224125 * max(0, X10 - 0)+ 8.32765 * max(0, X43 - 9.7619)- 0.301039 * max(0, X44 - 25.2381)- 0.694501 * max(0, 25.2381 -X44)- 0.0106345 * max(0,X6 - 11) + 1.2721 * max(0, 11 -X6) - 0.181544 * max(0, X3 - 4) * max(0, 11 -X6)- 0.548071 .|{X24 : a} * max(0, X10 - 0)- 0.637464 * max(0, X9 - 4)

Table 0.21-MARS model for “annual source energy,”,” with MBF=100, MI=2, and MOBN=2, based on Kung’s design

183.119 - 8.66819 * max (0, X43 - 22.4603) + 10.1056 * max (0, 22.4603 -X43) + 0.225443 * max (0, X44 - 10) * max (0, 22.4603 -X43) + 4.1687 .|{X1 : a}+ 65.9153 * max(0,X38 - 0.06)+ 12.8351 .|{X2 : c, d} .|{X1 : a} + 8.47895 * max(0, X43 - 9.7619) - 0.717305 * max(0, 25.2381 -X44)+ 1.65918 * max(0, 11 -X6) - 0.174279 * max(0,X3 - 4) * max(0, 11 -X6)- 0.699678 .|{X24 : a} * max(0, X10 - 0)

Table 0.22-MARS model for “annual source energy,” with MBF=50, MI=2, and MOBN=5, based on Kung’s design

237.296 - 5.90955 * max(0, X43 - 24.8413)+ 5.20763 * max(0, 24.8413 -X43)+ 0.207727 * max(0,X44 - 10) * max(0, 24.8413 -X43)+ 19.5053 .|{X1 : a} - 0.356096 * max(0,X6 - 0) .|{X1 : a}+ 105.368 .|{X2 : c, d} * max(0,X38 - 0.06)+ 5.69225 * max(0, X43 - 15.3175)- 0.838344 * max(0, 25.3968 -X44) - 0.644097 * max(0, X10 - 0)+ 7.54375 .|{X32 : a} .|{X1 :a} + 35.7464 * max(0, X37 - 0.06) .|{X1 : b} - 0.0938374 * max(0,X6 - 11)+ 0.30568 * max(0, 11 -X6) - 8.44004 .|{X2 : c, d}.|{X1 : b} - 1.22182 * max(0, X9 - 4) .|{X1 : a} - 2.40514 * max(0, X52 - 23.0952) .|{X1 : a} - 0.329855 * max(0, 23.0952 - X52) .|{X1 : a} + 0.224326 * max(0, 8 -X3) * max(0, 11 -X6)+ 0.00861674 * max(0, X43 - 35.1587) * max(0, 25.3968 -X44)+ 0.0373703 * max(0, 35.1587 -X43) * max(0, 25.3968 -X44)+ 0.42578 * max(0, 11 - X20) *.|{X1 : b} - 80.3546 * max(0, 0.12 - X35) .|{X1 : a} - 3.28651 .|{X33 : c, d} .|{X1 : a} + 0.952277 * max(0, 590.811 -X41) - 0.0897152 * max(0, X20 - 19) * max(0, 590.811 -X41) + 0.0506528 .|{X15 : b} * max(0, X41 - 590.811)+ 2.5437 * max(0, X37 - 0.06) * max(0, X10 - 0) + 8.47526 * max(0, X42- 24.4882) * max(0,X38 - 0.06) + 2.58423 * max(0, 24.4882 - X42) * max(0,X38 - 0.06) - 0.11454 * max(0, 15.9843 - X42) * max(0, 25.3968 -X44) + 0.0646042 * max(0, X40 - 24.9606) * max(0,X6 - 11) + 0.143007 * max(0, X43 - 68.4921)

.|{X1 : a} + 0.0682309 * max(0, 68.4921 -X43) .|{X1 : a} - 3.45309 * max(0, X4 - 4) * max(0,X38 - 0.06)

Table 0.23-MARS model for “annual source energy,” with MBF=100, MI=2, and MOBN=5, based on Kung’s design

252.139 - 4.81113 * max(0, X43 - 24.8413) + 5.4801 * max(0, 24.8413 -X43) + 0.157852 * max(0, X44 - 10) * max(0, 24.8413 -X43) + 12.3565 .|{X1 : a} - 0.388009 * max(0,X6 - 0) .|{X1 : a}+ 116.855 .|{X2 : c, d} *max(0, X38 - 0.06) + 4.64085 * max(0,X43 - 15.3175) - 0.633691 * max(0, 25.3968 -X44) - 0.773232 * max(0, X10 - 0) + 6.70435 .|{X32 : a} * (.|{X1 : a}) - 9.17655 .|{X2 : c, d}.|{X1 : b} - 1.07901 * max(0, X9 - 4) .|{X1 : a}+ 0.345791 * max(0, 8 -X3) * max(0, 11 - X6) + 3.6643 * max(0, X37 - 0.06) * max(0, X10 - 0)

Table 0.24-MARS model for “annual source energy,” with MBF=20, MI=2, and MOBN=5, based on Kung’s design

241.249 - 5.27931 * max(0, X43 - 24.8413) + 5.94607 * max(0, 24.8413 -X43) + 0.153901 * max(0, X44 - 10) * max(0, 24.8413 -X43) + 6.72709 .|{X1 : a} + 117.988 .|{X2 : c, d} * max(0, X38 - 0.06) + 5.12723 * max(0, X43 - 15.3175) - 0.663318 * max(0, 25.3968 -X44) - 0.526244 * max(0, X10 - 0) + 8.13895 .|{X32 : a} .|{X1 : a} + 55.292 * max(0, X37 - 0.06) .|{X1 : b} + 0.977851 * max(0, 11 -X6) - 9.18076.|{X2 : c, d}.|{X1 : b}

Table 0.25-MARS model for “annual source energy,” with MBF=20, MI=3, and MOBN=5, based on Kung’s design

241.249 - 5.27931 * max(0,X43 - 24.8413) + 5.94607 * max(0, 24.8413 -X43) + 0.153901 * max(0, X44 - 10) * max(0, 24.8413 -X43) + 6.72709 .|{X1 : a} + 117.988 .|{X2 : c, d} * max(0, X38 - 0.06) + 5.12723 * max(0, X43 - 15.3175) - 0.663318 * max(0, 25.3968 -X44) - 0.526244 * max(0, X10 - 0) + 8.13895 .|{X32 : a}.|{X1 : a} + 55.292 * max(0, X37 - 0.06) .|{X1 : b} + 0.977851 * max(0, 11 -X6) - 9.18076 .|{X2 : c, d}.|{X1 : b

Table 0.26-MARS model for “GWP,” with MBF=20, MI=1, and MOBN=2, based on Kung’s design

37600 + 2391.01 * max(0,X4 - 4) + 233.332 * max(0,X6 - 0) + 230.405 * max(0, X20 - 0) + 2.10451* max(0, X5 - 3000) - 4297.88.|{X23 : c}

Table 0.27-MARS model for “GWP,” with MBF=50, MI=1, and MOBN=2, based on Kung’s design

36170.1 + 2390.95 * max(0,X4 - 4) + 231.986 * max(0,X6 - 0) + 230.4 * max(0, X20 - 0) + 2.1164 * max(0,X5 - 3000)

Table 0.28-MARS model for “GWP,” with MBF=50, MI=2, and MOBN=10, based on Kung’s design

40754.3 + 2435.03 * max(0, X4 - 4) + 14.4942 * max(0, X20 - 0) * max(0,X6 - 0) + 46264.9 .|{X29 : a} * max(0,X35 - 0.06) - 0.309816 * max(0, 4000 -X5) * max(0,X6 - 0) + 1894.42 * max(0, X36 - 0.12) * max(0,X6 - 0)

Table 0.29-MARS model for “non-renewable energy,” with MBF=20, MI=1, and MOBN=2, based on Kung’s design

471237 + 139732 .|{X23 : c} + 20047.2 * max(0, X4 - 4) + 3039.41 * max(0,X6 - 0) + 2768.41 * max(0,X20 - 0) - 64279.2 .|{X23 : a} + 393504 * max(0, X35 - 0.06) - 47795.5 *.|{X14 : a}+ 42008.6 .|{X29 : a}+ 6649.33 * max(0,X3 - 4) + 17.3149 * max(0,X5 - 3000) + 286439 * max(0, X38 - 0.06) + 1580.58 * max(0, X19 - 0)

Table 0.30-MARS model for “non-renewable energy,” with MBF=50, MI=1, and MOBN=2, based on Kung’s design

624249 + 138247 .|{X23 : c} + 20048.8 * max(0,X4 - 4) + 3007.02 * max(0,X6 - 0) - 64279.2 .|{X23 : a} - 49203.8 .|{X14 : a} + 1119650 * max(0, X35 - 0.18) + 4039.84 * max(0,X20 - 11)

Table 0.31-MARS model for “non-renewable energy,” with MBF=100, MI=3, and MOBN=5, based on Kung’s design

728158 - 63522.8 .|{X23 : a } + 226037 * max(0, X27 - 5.96046e-008) .|{X23 : c}

Table 0.32-Treed regression model for “annual source energy,” by considering all variables in tree generation and at least 57 cases to split TN, based on Martinez’s design

TN #	Fitted Model
1	$[356.76753+148.85059\times X_{35}+153.59754\times X_{36}-8.75009\times X_{43}+3.32697\times X_{44}-1.32453\times X_{52}].\|\{X_{43}<=18.95\}$
2	$[264.21278-2.91498\times X_{4}+83.22190\times X_{36}+131.62426\times X_{38}+0.97969\times X_{44}].\|\{X_{43}>18.95\}.\|\{X_{43}<=47.96\}$
3	$[241.52891-0.28725\times X_{6}-0.74468\times X_{10}+37.87256\times X_{35}+80.68476\times X_{36}+57.90064\times X_{37}+90.27197\times X_{38}+0.47642\times X_{48}].\|\{X_{43}>18.95\}.\|\{X_{43}>47.96\}$

Table 0.33-Treed regression model for “annual source energy,” by considering all variables in tree generation and at least 58 cases to split TN, based on Martinez’s design

TN #	Fitted Model
1	$[231.66324+68.35021\times X_{35}+78.66452\times X_{36}+53.95842\times X_{37}+120.41028\times X_{38}-0.40318\times X_{43}+1.51893\times X_{44}].\|\{X_{28}:a\}$
2	$[309.46438-1.00621\times X_{10}+86.57597\times X_{36}+147.13991\times X_{38}-0.73429\times X_{43}].\|\{X_{28}:b\}$

Table 0.34-Treed regression model for “annual source energy,” by considering only the categorical variables in tree generation and at least 60 cases to split TN, based on Martinez’s design

TN #	Fitted Model
1	$[310.54855+133.77567\times X_{38}+1.20901\times X_{42}-6.38271\times X_{43}+3.45274\times X_{44}].\|\{X_{43}<=22.74\}$
2	$[267.18095-3.47451\times X_{4}+85.13819\times X_{36}+121.44777\times X_{38}+1.00874\times X_{44}].\|\{X_{43}>22.74\}.\|\{X_{43}<=47.96\}$
3	$[241.52891-0.28725\times X_{6}-0.74468\times X_{10}+37.87256\times X_{35}+80.68476\times X_{36}+57.90064\times X_{37}+90.27197\times X_{38}+0.47642\times X_{48}].\|\{X_{43}>22.74\}.\|\{X_{43}>47.96\}$

Table 0.35-Treed regression model for “GWP” without interaction, by considering only the categorical variables in tree generation and at least 60 cases to split TN, based on Martinez’s design

TN #	Fitted Model
1	$[9691.82630+2583.04564\times X_{4}+2.71547\times X_{5}+152.83489\times X_{6}+238.92693\times X_{10}+125.62589\times X_{19}+94.10201\times X_{20}+10406\times X_{37}+14402\times X_{37}].\|\{X_{1}:b\}$
2	$[11256+2840.17454\times X_{4}+2.21506\times X_{5}+136.31689\times X_{6}+284.79111\times X_{10}+28.18936\times X_{18}+158.78896\times X_{19}+147.47829\times X_{20}+11936\times X_{35}].\|\{X_{1}:a\}.\|\{X_{23}:a,c\}$
3	$[18024+2652.39468\times X_{4}+2.36551\times X_{5}+175.69947\times X_{6}+126.48855\times X_{19}+211.28508\times X_{20}+14633\times X_{36}].\|\{X_{1}:a\}.\|\{X_{23}:b\}$

Table 0.36-Treed regression model for "GWP" with interaction, by considering only the categorical variables in tree generation and at least 60 cases to split TN, based on Martinez's design

TN #	Fitted Model
1	[9438.32670+2584.75243×X_4+802.67201×stdX_4X_5+2.69204×X_5+157.19755×X_6 +238.77983×X_{10} +118.72113×X_{19} + 95.75040×X_{20} +12498×X_{37} +14558×X_{38}].|{X_1:b}
2	[11937+2880.18386×X_4+915.85765×stdX_4X_5+2.31334×X_5+160.06227×X_6+227.28705×X_{10} +129.21817×X_{19} +128.89090×X_{20}+10931×X_{35}+9809.71977×X_{36}].|{X_1:a}.|{X_{23}:a,c}
3	[20567+2399.06279×X_4+974.28132×stdX_4X_5+2.13880×X_5+156.80923×X_6+124.45447×X_{19} +224.57379×X_{20} +9477.56001×X_{36}+10074×X_{37}].|{X_1:a}.|{X_{23}:b}

Table 0.37-Treed regression model for "GWP," by considering all variables in tree generation and at least 69 cases to split TN, based on Martinez's design

TN #	Fitted Model
1	[37304 + 285.72632×X_6].|{X_4<=6}.|{X_5<=4500}
2	[39215 + 100.6277×X_6+206.86575×X_{10}+145.64022×X_{19}+165.01132×X_{20}].|{X_4<=6}.|{X_5>4500}
3	[40894+565.27675×X_3+247.16515×X_6+136.87828×X_{19}].|{X_4>6}.|{X_5<=3500}
4	[33331+2.14261×X_5+101.10425×X_6+108.16504×X_{19}+191.17158×X_{20}+7115.56249×X_{27}+19322×X_{38}] .|{X_4>6} .|{X_5>3500}

Table 0.38-Treed regression model for "GWP," by considering all variables in tree generation and at least 70 cases to split TN, based on Martinez's design

TN #	Fitted Model
1	[26938 + 2.08882×X_5+149.54011×X_6+251.43082×X_{10}+163.67100×X_{19}+133.32271×X_{20}].|{X_4<=6}
2	[23497+379.10049×X_3+4.23182×X_5+167.80047×X_6+168.57630×X_{19}+181.90505×X_{20}+26250×X_{38}].|{X_4>6} .|{X_5<=4500}
3	[49823+49823×X_6+49823×X_9+109.49680×X_{20}+14887×X_{35}].|{X_4>6}.|{X_5>4500}

Table 0.39-Treed regression model for "non-renewable energy," by considering all variables in tree generation and at least 72 cases to split TN, based on Martinez's design

TN #	Fitted Model
1	[324102 + 324102×X_3+33.93472×X_5+5300.82499×X_{10}+3095.57776×X_{10}].|{X_{23}:a,b}.|{X_4<=6}
2	[438960+30.60872×X_5+1981.66436×X_6+2249.77735×X_{20}+264301×X_{35}].|{X_{23}:a,b}.|{X_4>6}.|{X_{23}:a}
3	[668278+2910.53134×X_{20}+378506×X_{35}].|{X_{23}:a,b}.|{X_4>6}.|{X_{23}:b}
4	[415597+21241×X_4+25.40264×X_5+8158.28399×X_9+2116.28457×X_{19}+600585×X_{36}].|{X_{23}:c}.|{X_{25}:b}
5	[492734+6956.03026×X_3+30352×X_4+21.49444×X_5+1846.79116×X_{20}].|{ X_{23}:c}.|{X_{25}:a}

Table 0.40-Treed regression model for "non-renewable energy," by considering all variables in tree generation and at least 73 cases to split TN, based on Martinez's design

TN #	Fitted Model
1	[324102 + 8651.02767×X_3+33.93472×X_5+5300.82499×X_{10}+3095.57776×X_{19}].|{X_{23}:a,b}.|{X_4<=6}
2	[622043+1915.95214×X_6+2546.59271×X_{19}+2261.01881×X_{20}].|{ X_{23}:a,b}.|{X_4>6}
3	[485160+7580.9054×X_3+23949×X_4+32.05505×X_5+2450.79438×X_6-141666×X_{27}+428580×X_{36}].|{X_{23}:c}

Table 0.41-Treed regression model for “non-renewable energy” without interactions, by considering only the categorical variables in tree generation and at least 72 cases to split TN, based on Martinez’s design

TN #	Fitted Model
1	$[212957 + 25530\times X_4 + 27.40940\times X_5 + 1794.46810\times X_6 + 5611.81698\times X_8 + 2094.89585\times X_{10} + 1716.94724\times X_{19} + 2534.17295\times X_{20} + 170320\times X_{35}].\vert\{X_{23}:a\}$
2	$[381106 + 23906\times X_4 + 18.87820\times X_5 + 2628.57264\times X_6 + 2362.85181\times X_{19} + 2269.81855\times X_{20}].\vert\{X_{23}:b\}$
3	$[415597 + 21241\times X_4 + 25.40264\times X_5 + 8158.28399\times X_9 + 2116.28457\times X_{19} + 600585\times X_{36}].\vert\{X_{23}:c\}.\vert\{X_{25}:b\}$
4	$[492734 + 6956.03026\times X_3 + 30352\times X_4 + 21.49444\times X_5 + 1846.79116\times X_{20}].\vert\{X_{23}:c\}.\vert\{X_{25}:a\}$

Table 0.42-Treed regression model for “non-renewable energy” without interactions, by considering only the categorical variables in tree generation and at least 73 cases to split TN, based on Martinez’s design

TN #	Fitted Model
1	$[212957 + 25530\times X_4 + 27.40940\times X_5 + 1794.4681\times X_6 + 5611.81698\times X_8 + 2094.89585\times X_{10} + 1716.94724\times X_{19} + 2534.17295\times X_{20} + 170320\times X_{35}].\vert\{X_{23}:a\}$
2	$[381106 + 23906\times X_4 + 18.87820\times X_5 + 2628.57264\times X_6 + 2362.85181\times X_{19} + 2269.81855\times X_{20}].\vert\{ X_{23}:b\}$
3	$[485160 + 7580.90540\times X_3 + 23949\times X_4 + 32.05505\times X_5 + 2450.79438\times X_6 - 141666\times X_{27} + 428580\times X_{36}].\vert\{X_{23}:c\}$

Table 0.43-MARS model for “annual source energy,” with MBF=20, MI=2, and MOBN=10, based on Martinez’s design

273.86 - 0.168049 * max(0, X43 - 16.0111) + 3.49265 * max(0, 16.0111 - X43) + 0.433913 * max(0, X44 - 10) * max(0, 16.0111 - X43) + 243.565 * max(0, X38 - 0.18) - 57.862 * max(0, 0.18 - X38) + 84.9094 * max(0, X36 - 0.06) + 241.615 * max(0, X37 - 0.18) + 0.732467 * max(0, 11 - X6) + 0.0381782 * max(0, 57.0421 - X43) * max(0, X44 - 10) + 20.1595 .|{X1 : a} - 12.0653 .|{X2 : a, b} .|{X1 : a} - 1.38384 * max(0, X40 - 20.5539).|{X1 : a} - 1.35155 * max(0, 20.5539 - X40).|{X1 : a}

Table 0.44-MARS model for “annual source energy,” with MBF=100, MI=3, and MOBN=10, based on Martinez’s design

278.34 + 9.60849 * max(0, 16.0111 - X43) + 0.0494623 * max(0, 68.3232 - X43) * max(0, X44 - 10)

Table 0.45-MARS model for “annual source energy,” with MBF=20, MI=1, and MOBN=10, based on Martinez’s design

271.928 - 0.5024 * max(0, X43 - 16.0111) + 9.25181 * max(0, 16.0111 - X43) + 96.9446 * max(0, X38 - 0.06) + 0.934119 * max(0, X44 - 10) + 85.4636 * max(0, X36 - 0.06) + 0.800495 * max(0, 11 - X6) + 142.51 * max(0, X37 - 0.18) + 7.89006 .|{X1 : a} + 0.517627 * max(0, X43 - 68.3232) - 5.91222 .|{X2 : a} - 0.534547 * max(0, X10 – 0) + 39.9857 * max(0, X35 - 0.06)

Table 0.46-MARS model for “annual source energy,” with MBF=100, MI=2, and MOBN=10, based on Martinez’s design

280.729 + 0.798626 * max(0, X44 - 10) * max(0, 16.0111 - X43) + 0.0421451 * max(0, 68.3232 - X43) * max(0, X44 - 10)

Table 0.47-MARS model for response 1, with MBF=20, MI=2, and MOBN=2, based on the second design

304.199 - 0.379389 * max(0, X43 - 15.5083) + 0.960119 * max(0, X44 - 10) * max(0, 15.5083 - X43)

Table 0.48-MARS model for “annual source energy,” with MBF=20, MI=2, and MOBN=5, based on Martinez’s design

305.124 - 0.391286 * max(0, X43 - 14.9435) + 1.02074 * max(0, X44 - 10) * max(0, 14.9435 - X43)

Table 0.49-MARS model for “annual source energy,” with MBF=20, MI=1, and MOBN=5, based on Martinez’s design

243.262 - 12.5293 * max(0, X43 - 14.9435) + 14.3909 * max(0, 14.9435 - X43) + 105.067 * max(0, X38 - 0.06) + 0.960532 * max(0, X44 - 10) + 87.6676 * max(0, X36 - 0.06) + 0.873694 * max(0, 11 - X6) + 185.151 * max(0, X37 - 0.18) + 6.5822 .|{X1 : a} + 1.28512 * max(0, X43 - 31.4912) + 11.0087 * max(0, X43 - 11.4304)

Table 0.50-MARS model for “GWP,” with MBF=20, MI=2, and MOBN=10, based on Martinez’s design

37643.8 + 3109.37 * max (0, X4 - 4) - 2.61866 * max (0, 5000 - X5) + 4847.06 .|{X23 : b} + 148.181 *max(0,X6 – 0) + 86.2849 * max(0,X20 - 0) + 95.5486 * max(0, X19 -0) + 11142.9 * max(0, X35 - 0.06) + 149.399 * max(0, X10 - 0) + 777.901 * max(0, X36 - 0.06) * max(0, X20 - 0) - 0.399536 * max(0, X4 - 4) * max(0, 5000 - X5) - 2282.19 .|{X28 : b} .|{X32 : a}

Table 0.51-MARS model for “GWP,” with MBF=100, MI=1, and MOBN=2, based on Martinez’s design

38244.1 + 2817.93 * max (0, X4 - 4) - 2.57982 * max (0,5000 - X5)+ 4792.79 .|{X23 : b} + 151.901 * max (0, X6 - 0) + 156.731 * max(0, X20 - 0) - 1906.41.|{X32 : a} + 153.028 * max(0, X10 - 0) - 634.55 * max(0, 6 - X9) - 880.321 .|{X14 : a} + 13366 * max(0, X35 - 0.12) - 702.644 .|{X29 : b} + 872.839 .|{X25 : a} - 752.855 .|{X13 : a} - 612.545 .|{X28 : b} + 10841.4 * max(0, X36 - 0.12) + 125.55 * max(0, X19 - 8) + 1.2225 * max(0, X5 - 4000) + 8107.41 * max(0, X38 - 0.12)

Table 0.52-MARS model for “GWP,” with MBF=50, MI=1, and MOBN=2, based on Martinez’s design

38004.4 + 2808.89 * max(0, X4 - 4) - 2.73383 * max(0, 5000 - X5) + 4813.06 .|{X23 : b}+ 153.661 * max(0, X6 - 0) + 157.385 * max(0, X20 - 0) - 1962.99 .|{X32 : a} + 150.888 * max(0, X10 - 0) - 659.257 * max(0, 6 - X9) - 836.506 .|{X14 : a} + 5994.8 * max(0, X38 - 0.06) + 13705.7 * max(0, X35 - 0.12) - 734.12 .|{X29 : b} + 880.764 .|{X25 : a} - 723.561 .|{X13 : a} + 9675.22 * max(0, X36 - 0.12) + 125.634 * max(0, X19 - 8) + 1.09367 * max(0, X5 - 4000)

Table 0.53-MARS model for “GWP,” with MBF=20, MI=3, and MOBN=2, based on Martinez’s design

38414.8 + 2814.3 * max(0, X4 - 4) - 3.52409 * max(0, 5000 - X5)+ 4689.11 .|{X23 :b} + 138.068 * max(0, X6 - 0) + 82.4884 * max(0, X20 - 0) + 98.7532 * max(0, X19 - 0) +

11368.3 * max(0, X35 - 0.06) + 149.422 * max(0, X10 - 0) + 767.603 * max(0, X36 - 0.06) * max(0, X20 - 0) - 2226.48 .|{X28 : b} .|{X32 : a}

Table 0.54-MARS model for “GWP,” with MBF=50, MI=2, and MOBN=10, based on Martinez’s design

38919.3 + 3104.02 * max(0, X4 – 4) - 2.65953 * max(0, 5000 - X5) + 4789.08 .|{X23 : b } + 144.528 * max(0, X6 - 0) + 97.8939 * max(0, X19 - 0) + 143.644 * max(0, X10 - 0) + 897.34 * max(0, X36 - 0.06) * max(0, X20 - 0) - 0.391478 * max(0, X4 - 4) * max(0, 5000 - X5) - 2328.69 .|{X28 : b} .|{X32 : a} + 671.227 * max(0, X20 - 0) * max(0, X35 - 0.06)

Table 0.55-MARS model for “GWP,” with MBF=20, MI=1, and MOBN=2, based on Martinez’s design

39332 + 2806.42 * max (0, X4 - 4) - 3.58605 * max (0, 5000 - X5) + 4505.12 .|{X23 : b} + 154.769 * max(0, X6 - 0) + 160.36 * max(0, X20 - 0) - 2105.07 .|{X32 : a} + 104.456 * max(0, X19 - 0) + 135.192 * max(0, X10 - 0) - 727.286 * max(0, 6 - X9) + 16237.7 * max(0, X35 - 0.12)

Table 0.56-MARS model for “non-renewable energy,” with MBF=20, MI=2, and MOBN=2, based on Martinez’s design

554859 + 111871 .|{X23 : c} + 29234.8 * max(0, X4 – 4) - 87368 .|{X23 : a} + 23.9072 * max(0, X5 - 3000) + 1897.19 *max(0, X6 - 0) + 934.783 * max(0, X20 - 0) - 37722.1 .|{X14 : a} + 2011.32 * max(0, X19 -0) - 26501.5 .|{X32 : a} + 6945.85 * max(0, X3 - 4) .|{X23 : c} + 12618.6 * max(0, X36 - 0.06) * max(0, X20 - 0)+ 1835.53 * max(0, X10 - 0) - 6405.82 .|{X28 : b} * max(0, X4 - 4) + 253776 * max(0, X35 - 0.12) .|{X23 : a, b }

Table 0.57-MARS model for “non-renewable energy,” with MBF=20, MI=1, and MOBN=2, based on Martinez’s design

548361 + 130920 .|{X23 : c} + 25295.7 * max(0, X4 - 4) - 81993.4 .|{X23 : a} + 24.1823 * max(0, X5 - 3000) + 1837.05 * max(0, X6 - 0) + 2109.75 * max(0, X20 – 0) - 40302.2 .|{X14 : a} + 2006.88 * max(0, X19 - 0) - 26087.2 .|{X32 : a} + 2250.83 * max(0, X10 - 0) + 179603 * max(0, X36 - 0.06)

Table 0.58-MARS model for “non-renewable energy,” with MBF=100, MI=1, and MOBN=2, based on Martinez’s design

550587 + 130797 .|{X23 : c} + 25783.8 * max(0, X4 - 4) - 80657.6 .|{X23 : a} + 24.2797 * max(0, X5 - 3000) + 1696.15 * max(0, X6 - 0) + 2176.4 * max(0, X20 - 0) - 40003.5 .|{X14 : a} + 1988.4 * max(0, X19 - 0) - 25663.3 .|{X32 : a} + 2113.42 * max(0, X10 - 0) + 3166.23 * max(0, X3 - 4) - 15432.4 .|{X28 : b} + 255868 * max(0, X36 - 0.12)

Table 0.59-MARS model for “non-renewable energy,” with MBF=50, MI=1, and MOBN=2, based on Martinez’s design

518277 + 135226 .|{X23 : b} + 26658.9 * max(0, X4 - 4) - 85313.8 .|{X23 : a} + 25.4753 * max(0, X5 - 3000) + 860.089 * max(0, X6 - 0) + 2192.81 * max(0, X20 - 0) - 37435.1 * (.|{X14 : a}) + 2033.96 * max(0, X19 - 0) - 27243.1 .|{X32 : a} + 2057.37 * max(0, X10 - 0) + 6485.98 * max(0, X3 - 4) - 15921.9 .|{X28 : b} + 123377 * max(0, X37 - 0.06) - 13802.8 .|{X11 : b}+ 2402.75 * max(0, X6 - 19) + 3028.79 * max(0, X8 - 0) + 114320 * max(0, X35 - 0.06) - 32376.1 * max(0, X27 - 0.5) + 8804.52 .|{X25 : a} - 6021.14 .|{X29 : b} - 5160.07 * max(0, X3 - 8) + 50388.4 * max(0, X38 - 0.06) - 199126 * max(0, X35 - 0.18) + 4484.17 * max(0, X9 - 4) - 6609.83 * max(0, X9 - 8) + 215089 * max(0, X36 - 0.12)

Table 0.60-MARS model for “non-renewable energy,” with MBF=100, MI=3, and MOBN=5, based on Martinez’s design

555383 + 130058 .I{X23 : c} + 29901.3 * max(0, X4 - 4) – 86639.5 .I{X23 : a} + 24.5626 * max(0, X5 - 3000) + 1914.49 * max(0, X6 - 0) – 44459.7 * (.I{X14 : a}) + 2106.43 * max(0, X19 - 0) + 17916.5 * max(0, X36 - 0.06) * max(0, X20 - 0) + 2471.74 * max(0, X10 - 0) – 8026.42 .I{X28 : b} * max(0, X4 - 4)

Table 0.61-MARS model for “non-renewable energy,” with MBF=50, MI=3, and MOBN=5, based on Martinez’s design

563217 + 133827 .I{X23 : c} + 25845.2 * max(0, X4 - 4) - 89321.1 .I{X23 : a} + 23.2421 * max(0, X5 - 3000) + 1973.74 * max(0, X6 - 0) + 2282.83 * max(0, X20 - 0) - 46332.7 .I{X14 : a} + 1911.8 * max(0, X19 - 0)

REFERENCES

[1] A. Perujo and B. Ciuffo. The introduction of electric vehicles in the private fleet: Potential impact on the electric supply system and on the environment. A case study for the province of Milan, Italy. *Energy Policy 38(8),* pp. 4549-4561. 2010.

[2] N. H. Wong, D. W. Cheong, H. Yan, J. Soh, C. Ong and A. Sia. The effects of rooftop garden on energy consumption of a commercial building in singapore. *Energy Build. 35(4),* pp. 353-364. 2003.

[3] H. Castleton, V. Stovin, S. Beck and J. Davison. Green roofs; building energy savings and the potential for retrofit. *Energy Build. 42(10),* pp. 1582-1591. 2010.

[4] A. Brahma, Y. Guezennec and G. Rizzoni. Optimal energy management in series hybrid electric vehicles. Presented at American Control Conference, 2000. Proceedings of the 2000. 2000.

[5] A. Noth, W. Engel and R. Siegwart. Design of an ultra-lightweight autonomous solar airplane for continuous flight. Presented at Field and Service Robotics. 2006.

[6] I. Z. Bribián, A. A. Usón and S. Scarpellini. Life cycle assessment in buildings: State-of-the-art and simplified LCA methodology as a complement for building certification. *Build. Environ. 44(12),* pp. 2510-2520. 2009.

[7] *http://doe2.com/download/equest*

[8] *http://www.athenasmi.org*

[9] A. I. Forrester and A. J. Keane. Recent advances in surrogate-based optimization. *Prog. Aerospace Sci. 45(1),* pp. 50-79. 2009.

[10] V. C. Chen, K. Tsui, R. R. Barton and M. Meckesheimer. A review on design, modeling and applications of computer experiments. *IIE Transactions 38(4),* pp. 273-291. 2006.

[11] J. Sacks, W. J. Welch, T. J. Mitchell and H. P. Wynn. Design and analysis of computer experiments. *Statistical Science* pp. 409-423. 1989.

[12] D. Voss. Design and analysis of experiments. 1999.

[13] D. C. Montgomery. *Design and Analysis of Experiments* 2008.

[14] P. Kung. Multivariate modeling for A multiple stage, multiple objective green building framework. 2013.

[15] A. S. Hedayat, N. J. A. Sloane and J. Stufken. *Orthogonal Arrays: Theory and Applications* 2012.

[16] S. IM. The distribution of points in a cube and the approximate evaluation of integrals. *USSR Computational Mathematics and Mathematical Physics 7*pp. 86-112. 1976.

[17] M. D. McKay, R. J. Beckman and W. J. Conover. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics 42(1),* pp. 55-61. 2000.

[18] N. M. Martinez Cepeda. Global optimization of nonconvex piecewise linear regression splines. 2013.

[19] W. P. Alexander and S. D. Grimshaw. Treed regression. *Journal of Computational and Graphical Statistics 5(2),* pp. 156-175. 1996.

[20] L. Breiman, J. Friedman, C. J. Stone and R. A. Olshen. *Classification and Regression Trees* 1984.

[21] J. H. Friedman. Multivariate adaptive regression splines. *The Annals of Statistics* pp. 1-67. 1991.

[22] James J. Hirsch and Associates (JJH),
"the QUick Energy Simulation Tool," 2016.

[23] T. Maile, M. Fischer and V. Bazjanac. Building energy performance simulation tools-a life-cycle and interoperable perspective. *Center for Integrated Facility Engineering (CIFE) Working Paper 107*pp. 1-49. 2007.

[24] https://www.salford-systems.com.

[25] T. Hong, S. Chou and T. Bong. Building simulation: An overview of developments and information sources. *Build. Environ. 35(4),* pp. 347-361. 2000.

[26] I. Z. Bribián, A. V. Capilla and A. A. Usón. Life cycle assessment of building materials: Comparative analysis of energy and environmental impacts and evaluation of the eco-efficiency improvement potential. *Build. Environ. 46(5),* pp. 1133-1140. 2011.

[27] T. Wooley, S. Kimmins, P. Harrison and R. Harrison. *Green Building Handbook: A Guide to Building Products and their Impact on the Environment* 1997.

[28] W. Wang, R. Zmeureanu and H. Rivard. Applying multi-objective genetic algorithms in green building design optimization. *Build. Environ. 40(11),* pp. 1512-1525. 2005.

[29] X. Zhai, R. Wang, Y. Dai, J. Wu, Y. Xu and Q. Ma. Solar integrated energy system for a green building. *Energy Build. 39(8),* pp. 985-993. 2007.

[30] M. Asif, T. Muneer and R. Kelley. Life cycle assessment: A case study of a dwelling home in scotland. *Build. Environ. 42(3),* pp. 1391-1394. 2007.

[31] https://www.isixsigma.com/tools-templates/design-of-experiments-doe/design-experiments-%E2%90%93-primer/.

[32] https://en.wikipedia.org/wiki/Sobol_sequence.

[33] https://en.wikipedia.org/wiki/Latin_hypercube_sampling.

[34] W. Loh. Classification and regression trees. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery 1(1),* pp. 14-23. 2011.

[35] http://www.sas.com/en_us/home.html.

[36] http://neilsloane.com/oadir/.

[37] https://www.rstudio.com/.

[38] H. K. Shah, D. C. Montgomery and W. M. Carlyle. Response surface modeling and optimization in multiresponse experiments using seemingly unrelated regressions. *Quality Engineering 16(3),* pp. 387-397. 2004.

Biographical Information

Marjan Sayadi holds a Bachelor of Science (B.S.) in Statistics from the University of Tehran, Iran. She graduated from her B.S in 2006. She is a recipient of the fellowship for her studies during B.S. Marjan holds a Master of Science in Industrial Engineering from the University of Texas at Arlington. She graduated with a GPA of 4.0 in 2016. She is a recipient of the Society of Iranian-American Women for Education scholarship in 2015. Marjan is a member of Tau Beta Pi engineering honor society since 2016, and held a position of Graduate Teaching Assistant for three semesters in the department of Industrial Engineering.