BAYESIAN BASED PARAMETER IDENTIFICATION FOR BUILDING ENERGY MODELS

by

YOONSUK KANG

B.A., Architectural Engineering, Dong-A University, Korea, 2007

M.S., Architectural Engineering, Dong-A University, Korea, 2009

A thesis submitted to the

Faculty of the Graduate School of the

University of Colorado in partial fulfillment

Of the requirement for the degree of

Doctor of Philosophy

Department of Civil, Environmental, and Architectural Engineering

2014

This thesis entitled:

Bayesian based Parameter Identification for Building Energy Models

written by Yoonsuk Kang

has been approved for the Department of Civil, Environmental, and Architectural Engineering

Moncef Krarti, Ph.D (Advisor)

Michael J. Brandemuehl, Ph.D

John Zhai, Ph.D

Walter Beamer IV, Ph.D

Wisit Kumphai, Ph.D (CBRE)

Date

The final copy of this thesis has been examined by the signatories and we find that both the content and the form meet acceptable presentation standards of scholarly work in the above mentioned discipline.

Kang, Yoonsuk (Ph.D., Architectural Engineering)

Bayesian based Parameter Identification for Building Energy Models

A thesis directed by Professor Moncef Krarti, Ph.D., PE, LEED®AP

In this research work, a series of sensitivity analyses were performed to validate the proposed Bayesian approach to identify unknown parameters in building energy models. The proposed Bayesian approach mainly consisted of creating a Gaussian process emulator to sample the posterior distribution. Sensitivity case studies were carried out to investigate followings: appropriate sampling numbers, size of Gaussian process, observation noise, continuous/discrete variables situation. Validation on the proposed approach was done with closed loop results (one RC model and two DOE2.2 models) as well as three actual buildings (two commercial buildings and one residential building). The result showed success of identifying unknown parameters by higher occurrences on target values. Moreover, the proposed approach was tested in actual buildings and shown to calibrate the building energy models with unknown parameters still inside.

As an application of the proposed Bayesian approach, development of identification of Energy Conservative Measures (ECMs) were carried out. The proposed approach succeeded in identifying the appropriate ECMs with uncertainties in budgets, initial costs, and actual performance of the ECMs. Furthermore, comparison studies between other linear models and traditional Bayesian approach have been carried out to demonstrate the characteristic of the proposed approach to other methods. Also, this study has validated the possibility of utilizing the simplified approach in a future study.

ACKNOWLEDGEMENT

First and foremost, I would like to express the greatest gratitude to my family who has supported me through my long and endless journey towards the degree. Especially my parents (Yeonsil Kang, Yeongsuk Jeung), I thank them for their valuable advice and support.

I would also like to express the deepest thanks to my advisor Dr. Moncef Krarti for his guidance and support during my study in University of Colorado at Boulder. His knowledge, patience, and encouragement have given me invaluable guidance.

CONTENTS

List of Tables

List of Figures

CHAPTER 1: Introduction

1.1. Background

Currently, the building sector consumes 40% of all energy and is responsible for 30% of global annual greenhouse gas emissions. Within the overall building energy consumption, nearly 80% of energy consumption is dedicated to the operational phase of buildings (e.g. HVAC, lighting, appliances and other applications) (Sustainable Buildings & Climate Initiative, 2009). Considering the building has a longer life cycle, the energy consumption is a significant factor in the energy sectors.

The essential initial step to have energy consumption reduced from a building is to perform a building energy analysis. The building energy analysis is a fundamental procedure to determine the current condition of the building. The information obtained from the building energy analysis can be used toward energy retrofits and maybe used for the evaluations of potential improvement that can be applied. In order to achieve accurate estimation of further energy improvement, precise energy analysis is required. It is crucial that the building energy model in the building energy analysis mimics the current situation of the building. Misrepresentation of the building can lead to a faulty energy prediction in potential energy savings.

Normally, the building energy analysis requires a calibration procedure to represent the actual building. The calibration procedure is a procedure of parameter iterations that uses actual data and alters the parameter of the computer energy model to behave like an actual building. Actual data can come from energy usage from utility bills or experimental data logs (such as indoor temperature, relative humidity, lighting density, boiler efficiency, etc.) from the actual building site. By using this data, a building energy modeler can alter the building energy model's input values to be more appropriate and therefore achieve closer modeling of actual building behavior. However, even when a blueprint for a building or information from a field building audit is available, there are still uncertainties in the building energy analysis. This occurs for various reasons (e.g., deterioration of the material, unspecified properties in the building specifications, irregular building schedule, etc.). These uncertainties in parameters usually interrupts the identification of the true values. Therefore a parameter identification technique is required in the building energy analysis.

The parameter identification technique can be understood as identifying unknown parameters that can occur in the estimation of multiple-equation models, an unknown function model, or a sophisticated computer-based simulation model. The parameter identification technique can also be referred to as an “inverse problem,” in the sense that it inverts the process of solving the differential equations. In the field of engineering, there are various parameter identification techniques and various applications that require these techniques. In building energy analysis, the unknown parameter identification technique has been recognized as an essential part. Parameter identification can lack a certain amount of accuracy because of various uncertainties involved in the energy modeling process. These numerous uncertainties can be derived from obtaining and utilizing the building data, uncertainties from building energy modeler’s modeling capability, etc. Thus, quantifying uncertainties is crucial to achieving an accurate calibration, if left unattended this can lead to skewed parameter identification and mis-prediction of conditions regarding existing or new construction. In addition to the uncertainties in actual building parameters, uncertainties in the quantity of unknown parameters can impact the building modeling process. Consequently, various methods are applied in order to consider these uncertainties in building modeling. Traditionally, building energy experts utilized their knowledge and expertise to handle these uncertainties. Another method is to apply computerized methods such as optimization methods to handle the parameter identifications. However, optimization methods are usually deterministic and can lead to a faulty result of parameter identification. An alternative method to replacing the deterministic computerized methods would be applying statistical and probabilistic methods to account for the uncertainties within the building energy models. One representative method of apply statistical and probabilistic methods will be applying the Bayesian approach. Heo's study (Heo, 2011) utilized the Bayesian approach to quantify uncertainties and assist in the decision making process of building energy retrofits. In the parameter identification method, there remains numerous uncertainties within the building energy models. The Bayesian approach can quantify these uncertainties and therefore mitigate these uncertainties in the building parameter identification.

1.2. Justification of the research

Among various parameter identification techniques, applying an optimization method is known as a general approach to this situation. Other approaches include applying statistical and probabilistic methods to identify unknown parameters. Probabilistic analysis can be implemented by applying Bayesian theory. Boujelben (Boujelben, 2012) compared various optimization approaches to the Bayesian approach in building energy analysis. The results suggested that the Bayesian approach was the most efficient in computational time compared to a genetic algorithm, particle swarm optimization, and brute-force. In terms of accuracy, the Bayesian approach seems to be the best approach for parameter identification. Another benefit of using a Bayesian approach is that it can also quantify uncertainties within the identification process. This is important because parameters are sometimes difficult to identify during the modeling process. Experts normally use their experience and expertise to estimate these parameters and apply this knowledge to the modeling process. However, in probabilistic analysis this process can be quantified. Capturing this expertise and estimations about each parameter as a prior belief and applying it to a Bayesian methodology can solve this issue. Whether the prior belief of certain parameters is weak or strong, quantifying the prior belief will have a strong influence on identifying unknown parameters in the modeling process.

When quantifying uncertainties in building energy models in building energy simulations, using the Bayesian approach is required for the following reasons:

- Uncertainties in obtaining physical or behavioral parameters are higher since specification values from building blueprints can contain depreciations. Even with values obtained from a field audit can contain uncertainties in measurement errors. Therefore, quantifying these uncertainties will increase overall accuracy of the energy model.
- Calibration of the energy model requires effort, time, and expertise to perform accurate modeling. Building a model that can reduce these is crucial to determine the precise model.
- After calibration, uncertainties in ECMs still reside in the ECM selections. Quantifying these to be reflected in the selection can achieve more precise results.

1.3. Research objectives

The goal of this study is to develop a Bayesian-based parameter identification method that can be used in applications for building energy analysis. In order to achieve this goal, the approach has to: a) quantify uncertainties in user input, b) obtain valid and accurate identification of unknown parameters, and c) decrease computational burden (speed of computational time). Along with addressing these items, this study will also develop a framework to support an application that can be utilized in building energy analysis. To validate this framework, case studies and testing will be done with both a resistor-capacitor (RC) model and DOE2.2 model of a building. After the parameter identifications, this study will develop risk quantified ECMs selection procedures to select ECMs with various scenarios to achieve an overall energy modeling process. To verify this emulator's validity with other regression methodology such as linear models, this study will perform a comparison study with other methods. Due to difficulties of modeling a building with complex geometry, this study will test our approach to building with simplified geometry and determine the differences in parameter identifications due to its complication of geometry. Lastly, this study will find out the amount of computing effort that will be reduced by using our approach compared with the traditional Bayesian approach.

Specific objectives of the implemented work in this thesis include:

1) Development of an efficient Bayesian approach that will alleviate the CPU burden and process parameter identification to be used by individuals without engineering backgrounds. Since the traditional optimization and the Bayesian approach can consume a lot of time to solve parameter identification problems, the proposed approach reduces computing burden as well as the effort that is required to do energy modeling.
2) After development of the Bayesian approach, this study will validate the approach with fully extensive examination. Using a virtual energy model as well as actual building models, it is necessary to verify the approach with thorough evaluations.
3) Once the development and verification process is complete, it is essential to expand the developed approach to further applications. One example will be utilizing the approach in search of appropriate ECMs. Since ECMs contain a lot of uncertainties (such as uncertainties in initial cost or actual performance of ECMs), it is crucial to determine

the applicability of using the proposed approach in ECMs evaluation.

4) Comparison between the other regression models and the developed approach is important since using the Bayesian model can be less effective than that of simpler linear models. Not only a comparison study to other linear models, the developed approach has to be validated against traditional Bayesian methodology. The result of these comparison studies can show the effectiveness of the developed approach to the traditional approach.

1.4. Outline of the thesis

The research will provide a comprehensive analysis of the Bayesian approach combined with building energy models. The following outline briefly describes the research provided in each of the seven chapters.

Chapter 1 presents an overview of the scope, justification and main objectives of the research.

Chapter 2 introduces the literature review of current and past studies regarding optimization and Bayesian approaches in different fields.

Chapter 3 discusses about how a Bayesian approach would fit in building energy models. Additionally, this chapter will introduce the methodology of our technique compared to the traditional Bayesian approach along with some preliminary results.

Chapter 4 provides preliminary results of the proposed Bayesian approach using various hypothetical building models. This chapter will present details about the role of observation errors and assumptions that is required by Bayesian approach.

Chapter 5 reviews the validity of our proposed approach to actual buildings. Three buildings were selected to do case studies and provide results.

Chapter 6 describes applications of utilizing our proposed Bayesian approach such as selecting ECMs. This chapter also describes the comparisons between other emulator methodologies such as linear models. Potential validation of using simplified modeling methods instead of complex geometry building is also introduced in this chapter. The result of comparison between traditional Bayesian approaches is narrated as well.

Chapter 7 summarizes the main findings of the research work as well as recommendations for future work.

CHAPTER 2: Literature review

2.1. Optimization methods in parameter identification

Parameter identification is an estimation of unknown parameters within multiple equations or functions where those functions or equations have parameters in common. There are numerous methodologies available in various fields. Some representative approaches to utilizing this methodology come from applying optimization methods (e.g., genetic algorithms (GA), particle swarm optimization (PSO), fuzzy logic, neural network (NN), etc.).

Genetic algorithms (GA) are a family of computational models inspired by evolution (Whitley, 1994). This methodology converts a potential solution to a certain problem on a simple gene-like data structure and applies recombination structures in such a way as to preserve critical information. GA can also often be viewed as function optimizers despite being applied to quite a broad range. In the building sector, a study on GA for VAV air-conditioning system was done (Wang, et al., 2000). This study proposed a control strategy using a system approach based on predicting the responses of the overall environment and energy performance to change VAV system control settings by using a GA to solve an online optimization problem with multiple parameters. In another study (Caldas, et al., 2003) GA was utilized to solve multi-objective optimization problems in determining the size and placement of windows and the composition of building walls, the generation of building form, and the design and operation of HVAC systems. Wang (Wang, et al., 2006a) developed a genetic algorithm estimator to estimate the lumped internal thermal parameters of a building thermal network model using operation data collected from site monitoring. Additionally, Wang (Wang, et al., 2006b) developed a simplified building thermal model to represent physical details based on frequency characteristic analysis, since it is very difficult to obtain such detailed physical properties. The study developed GA estimators to identify parameters of a thermal network of a lumped thermal mass by using operation data. Xu (Xu, et al., 2007) presented a methodology for parameter optimization of 3R2C (composed of three resistance and two capacitances) thermal network building envelopes based on frequency domain regression using GA. A GA estimator was developed to optimize the parameters of the simplified model, allowing the frequency responses of the simplified model to match the actual heat transfer through the building envelope. Xu (Xu, et al., 2009) proposed a model-based optimal ventilation control strategy for multi-zone VAV air-conditioning systems aiming at optimizing the total fresh air flow rate without compromising thermal comfort, indoor air quality and total energy consumption. GA was used for optimizing the temperature set point of critical zones in the optimization process. Another study examined optimal VAV system control strategies (Mossolly, et al., 2009). One of the strategies adjusted the fresh air supply rate and the supply air temperature to maintain the temperature set point in each zone while assuring indoor air quality. Another strategy controlled the fresh air rate and the supply air temperature to maintain acceptable thermal comfort and IAQ in each zone. GA was used to solve the optimization problem for each control strategy based on the cost of energy consumption and constraints by system and thermal space transient models. Recently, a study was done on application of and GA (Genetic Algorithm) techniques to estimate oil demand in Iran based on socio-economic indicators (Assareh, et al., 2010). The models were developed in two forms (exponential and linear) and applied to forecast oil demand in Iran.

Particle swarm optimization (PSO) has roots in two main component methodologies. Perhaps more obvious are its ties to artificial life (A-life) in general, and to bird flocking, fish schooling, and swarming theory in particular (Kennedy, et al., 1995). It is also related, however, to evolutionary computation, and has ties to both genetic algorithms and evolutionary programming. Liu (Liu, et al., 2008) conducted a study on PSO, an intelligent computational method based on stochastic search, and it was shown to be a versatile and efficient tool for verifying the effectiveness for identification of permanent magnet synchronous motors model parameters. Zhang (Zhang, et al., 2009) conducted a study about PSO to identify the heat and moisture system parameters during delay in air-conditioned rooms by sampling input and output data. The study used PSO integrated with least square (LS) to improve least squares. Li (Li, et al., 2010) investigated the feasibility of using Least Squares Support vector regression (LS-SVR) to forecast building cooling load. Due to the importance of parameters optimization in a LS-SVR model, particle swarm optimization (PSO) was used to optimize the model parameters. The experiment results showed that PSO can quickly obtain the optimal parameters that satisfy precision requirements with a simple calculation, which solves the problem of complex calculation and empiricism in conventional methods.

Another popular optimization method is called neural network (NN). The term neural network was traditionally used to refer to a network or circuit of biological neurons (Hopfield, 1982). The modern usage of the term often refers to artificial neural networks, which are composed of artificial neurons or nodes. Mechaqrane (Mechaqrane, et al., 2004) proposed a neural network auto regressive with exogenous input (NNARX) model to predict the indoor temperature of a residential building. Ekici (Ekici, et al., 2009) used NN to predict buildings energy needs benefitting from orientation, insulation thickness and transparency ratio. The numerical applications were carried out with a finite difference approach for brick walls with and without insulation for transient state one-dimensional heat conduction. A computer program written in FORTRAN was used for the calculations of energy demand and the ANN toolbox of MATLAB was used for predictions. The energy consumption of a heating, ventilating and air conditioning (HVAC) system was optimized by using a data-driven approach (Kusiak, et al., 2012). This was done by utilizing the concept of a dynamic NN by building predictive models with controllable and uncontrollable input and output parameters.

There are also some studies that combine these optimization methods with other methods. Some of the studies utilize GA with fuzzy logic, mostly with fuzzy logic controllers. Fuzzy logic is a form of many-valued logic or probabilistic logic; it deals with reasoning that is approximate rather than fixed and exact. Compared to traditional binary sets (where parameters may take on true or false values) fuzzy logic parameters may have a true value that ranges in degree between 0 and 1. Fuzzy logic has been extended to handle the concept of partial truth, where the true value may range between completely true and completely false (Novak, et al., 1999). A study on developing smartly tuned fuzzy logic controllers dedicated to the control of HVAC systems concerning energy performance and indoor comfort requirements was done by Alcala (Alcala, et al., 2003). To solve some of the issues regarding the need to consider multiple criteria (which enlarges the solution search space) and to address the long computation time that some models require when assessing the accuracy of each individual, a genetic tuning strategy considering an efficient multi-criteria approach was proposed along with production of fuzzy logic controllers. A new experimental procedure was proposed based on a genetic algorithm for analyzing the exterior wall surface heat transfer process and deducing exterior wall surface heat transfer coefficients under actual conditions (Zhang, et al., 2004). Tests were carried out under a wide range of conditions, and the heat transfer coefficient was found to vary from 14.315 to 24.412 W/m2 K, while wind speed ranged from 1.04 to 7.36 m/s, and correlation was obtained in terms of the wind speed. Alcala (Alcala, et al., 2005) developed a genetic optimization process considering an efficient approach to perform rule weight derivation and rule selection for the use of weighted linguistic fuzzy rules in combination with a rule selection process to develop accurate fuzzy logic controllers dedicated to the intelligent control of HVAC systems concerning energy performance and indoor comfort requirements. A study (Lu, et al., 2005) presented a practical method to optimize in-building section of centralized HVAC systems that consists of indoor air loops and chilled water loops. An adaptive neuro-fuzzy inference system (ANFIS) was employed to model duct and pipe networks and obtain optimal differential pressure (DP) set points based on limited sensor information. Mix-integer nonlinear constraint optimization of system energy was formulated and solved by a modified genetic algorithm. Yan (Yan, et al., 2008) proposed an adaptive optimal control model that included the optimal control model, parameter identification and an optimization algorithm for a building cooling source system in a publishing house located in Changsha, China. First, a penalty function was constructed to transform the constrained optimization problem into an unconstrained optimization problem, and then, a fuzzy self-tuning forgetting factor method was developed for parameter identification. Finally, the genetic algorithm (GA) was used to determine optimal values for discrete and continuous parameters.

Additional studies have also utilized a combination of optimization methods. Chow (Chow, et al., 2002) introduced a concept of integrating NN and GA in the optimal control of an absorption chiller system. Based on a commercial absorption unit, NN was used to model system characteristics and GA was used as a global optimization tool.

Other studies have conducted a research to compare these optimization methods. Coelho (Coehlo, et al., 2005) studied PSO as a new method to design a model-based predictive greenhouse air temperature controller subject to restrictions and compared it with a GA and a sequential quadratic programming method. The results indicate better efficiency for the PSO. Wang (Wang, et al., 2008) provided a framework for categorizing the main supervisory and optimal control methods and optimization techniques developed and/or utilized in the HVAC field. A comprehensive overall review of the state of the art of the research and development, as well as application of supervisory and optimal control, in HVAC systems was also presented in the study.

2.2. Bayesian approach for parameter identification and calibration

Bayesian refers to methods in probability and statistics named after Thomas Bayes (c. 1702–61), and in particular methods related to statistical inference. Tse (Tse, et al., 1975) studied the problem of using output data to identify a constant, multivariable, stochastic linear system that has an unknown dimension, system matrices and noise covariance. The study used a representation for the output process, in which system matrices are chosen in a certain canonical form. Matsumoto (Matsumoto, et al., 1999) proposed a hierarchical Bayesian approach to nonlinear time series prediction problems and tested against two examples: (i) chaotic time series, and (ii) building air-conditioning load prediction. Kennedy (Kennedy, et al., 2001) proposed a consideration on prediction and uncertainty analysis for systems that are approximated using complex mathematical models. They presented a Bayesian calibration technique that improves on this traditional approach in two respects. First, the predictions allow for all sources of uncertainty, including the remaining uncertainty over the fitted parameters. Second, they attempt to correct for any inadequacy of the model that is revealed by a discrepancy between the observed data and the model predictions from even the best-fitting parameter values. A Bayesian network approach has been developed that can compare different building designs by estimating the effects of the thermal indoor environment on the mental performance of office workers (Jensen, et al., 2009). It focuses on the effects of temperature on mental performance and not on other indoor climate factors. This study verified that a Bayesian network provides a reliable platform when using probabilities for modeling complexity while estimating the effect of indoor climate factors on human beings, due to the different ways in which humans are affected by the indoor climate. Tarlow (Tarlow, et al., 2009) introduced a framework and proof of concept for estimating building energy consumption that probabilistically combines a model of building physics with observed occupancy and detailed operations data, which automatic learning for a physically plausible model of the energy consumption. Najafi (Najafi, et al., 2010) applied a Bayesian updating approach that is systematic in managing and accounting for most forms of the model and data errors that address cases where models are imperfect and data are variable, uncertain, and can contain error. This study demonstrated this approach by detecting faults in commercial building air handling units. Additionally, Najafi (Najafi, et al., 2012) developed diagnostic algorithms for air handling units that can address such constraints more effectively by systematically employing machine-learning techniques. The proposed algorithms were based on analyzing the observed behavior of the system and comparing it with a set of behavioral patterns generated based on various faulty conditions. Heo (Heo, et al., 2011) conducted a study on presenting a risk analysis method based on Bayesian calibration of building energy models. This study demonstrates the applicability of the proposed methodology to support energy saving contracts in the context of the ESCO industry. The study also conducted a case study that illustrates the importance of quantifying relative risks by comparing the probabilistic outputs derived from the Bayesian approach to standard practices endorsed by the International Performance Measurement and Verification Protocol and ASHRAE guideline 14. Another study (Heo, et al., 2012) presented a scalable, probabilistic methodology that can support large scale investments in energy retrofit of buildings while accounting for uncertainty. This study focused on Bayesian calibration of a normative energy model based on a normative CEN-ISO standard. Boujelben (Boujelben, 2012) presented a study of some optimization routines to improve the calibration of energy models for residential buildings using a

JAVA-Perl tool developed at the University of Colorado at Boulder. The first part deals with implementing two popular artificial intelligence algorithms: Genetic Algorithm and Particle Swarm Optimization. The latter was improved to fit a discrete space search. Based on Jumping Frogs Optimization, a new approach was developed to implement the Discrete Particle Swarm Optimization. The second part consisted of implementing a probabilistic approach based on Bayesian theory. The new approach introduced prior makes a calibration using Monte Carlo and Markov Chains methods. Lauret (Lauret, et al., 2013) proposed a Bayesian approach in the realm of solar radiation modeling. Manfren (Manfren, et al., 2013) studied an approach that is oriented to systems and able to effectively integrate field measured data and computer simulations for calibration in the modeling process and has the potential to revolutionize the way buildings are designed and operated, and to stimulate the development of new technologies and solutions in the field.

CHAPTER 3: Implementation of Bayesian parameter identification

3.1. Bayesian approach

In a normal optimization approach with unknown parameter identification, the main question is how to find the best suitable combination of each unknown parameter's value, such that it will make the observed data as close as possible to the reference data (in our case, utility data for electricity and/or gas). The term combination in this case combines all possible values for unknown parameters that require exploring a very large search space. This also indicates that the user has no background knowledge and expertise of these parameters. Therefore, users have no control over these optimization methods. And the result of optimization method is usually deterministic and contains a single set of values for unknown parameters. This is unreasonable for unknown parameter identification because occasionally users do not possess any knowledge about the unknown parameters and parameter identification usually derives solely from reference data. Because of this, in some cases parameter identification thru optimization methods can result in poor outcomes.

When performing energy analysis on an existing building or new construction, experts attempt to identify unknown parameters so as to place these unknown parameters in appropriate slots or fill in blanks in thermal equations or simulation tools. Some parameters are difficult to identify due to various reasons (e.g., deterioration of the material or system, unspecified in the blueprint, etc.). In these situations, experts usually use reference data or expertise to estimate these parameters. As such, unlike a normal optimization method, users do have some control over parameter identification during practical application. Reference data or expertise to estimate these parameters can be referred as prior belief. Prior belief means that whether the belief is weak or strong, there are some beliefs or background knowledge about the specified parameter. This is true for all energy analysis. Without prior belief, it is almost impossible for users to perform energy analysis. It would be advantageous to quantify and take into account this prior belief for parameter identification. The size of the search space can thus also be diminished and allow for shorter computational times. It would also be a practical approach and allow for accurate identification of appropriate parameters.

A Bayesian approach offers such a methodology. Bayesian parameter identification utilizes a statistical approach to make conclusions about phenomenon based on observed data, in our case, utility data. Bayesian methods utilize prior distribution (prior belief of the user) and observed data (utility bill) and calculates a posterior distribution, which is a conditional probability distribution of unknown parameters given observed data, and therefore results in a more practical methodology.

The goal of a Bayesian approach is to estimate the posterior distribution over a prior belief $p(\theta)$ and the data-likelihood of y given θ In energy model θ can be represented as a parameter that we wish to identify and $p(\theta)$ can be the user's certainty or confidence level in the parameter. And y can be a utility data that we wish to calibrate the model with. Using Bayes rule, we can evaluate the posterior distribution with:

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)} \tag{3.1}$$

For our study, there is no need to evaluate the denominator $p(y)$ since this is independent of θ. Therefore, we can ignore any constant of proportionality and rewrite this equation as

$$p(\theta|y) \propto p(y|\theta)\, p(\theta) \tag{3.2}$$

Thus, the posterior distribution can be estimated by a simple product of the likelihood and the prior belief. Basically a Bayesian approach to parameter estimation is to treat an unknown parameter as a random variable. For this random variable, a prior probability is assigned and the likelihood for specified variable is calculated and derives the posterior distribution. As shown in above equation, likelihood function $p(y|\theta)$ defines the closeness of the simulated data or emulated data for given parameters compared to the observed data.

In this thesis, the proposed unknown parameter identification follows a Bayesian approach that was proposed by Kennedy and O'hagan (Kennedy, et al., 2001). This study takes $\eta(x, t)$ to denote a simulator output given an input vector (x, t) where x indicates parameters that are known and controllable variables and t indicates those unknown parameters that need to be identified. For a suitable choice, such as t = θ, $\eta(x, \theta)$ simulates a physical system $\zeta(x)$. And physical system $\zeta(x)$ doesn't depend on θ. Therefore, various settings for x, and observations y can be made for a physical system,

$$y(x_i) = \zeta(x_i) + \epsilon(x_i), \qquad i = 1, \dots, n \tag{3.3}$$

where, $\epsilon(x_i)$ accounts for the observation error. If this physical system $\zeta(x)$ is replaced by simulation output $\eta(x, \theta)$, then the above equation can be changed into

$$y(x_i) = \eta(x, \theta) + \delta(x_i) + \epsilon(x_i), \qquad i = 1, \dots, n \tag{3.4}$$

where, $\delta(x_i)$ accounts for discrepancy between the simulation output $\eta(x, \theta)$ and physical system $\zeta(x_i)$. As shown in the above equation, the observation can be a sum of simulator output and discrepancy between simulation output and observation error.

In Bayesian data analysis, one way to apply a model to data is to find the maximum a posteriori (MAP) parameter values. The MAP can be used to obtain an estimate for an unobserved quantity on the basis of empirical data. The goal is to find the parameter estimates that maximize the posterior probability of the parameters given the observation data. This can be described as:

$$\theta_{MAP} = \arg_{\theta} \max p(\theta|y) \tag{3.5}$$

$$= \arg_\theta \max \frac{p(y|\theta)p(\theta)}{p(y)}$$

$$= \arg_\theta \max p(y|\theta)p(\theta)$$

In the study summarized in this paper, a Bayesian methodology was used to identify unknown parameters for a detailed building energy model. General energy model requires inputs that are unknown to general users because of the difficulties to identify the appropriate value. Our study is more focused on identification of unknown parameters with certain user's uncertainty in those variables. To eliminate the uncertainties regarding each variable, this study has quantified a prior belief. The intention of this study is not only to evaluate the generalized energy model, but also to evaluate the whole building simulation tool in order to validate the Bayesian approach.

The following unknown parameter identification process (Figure 3-1) was used in our study. Firstly, this study utilizes a user's belief for each variable and quantifies this uncertainty as a prior distribution. In order to better quantify the prior distribution for each variable, a user's belief (confidence interval; CI) was used. Secondly, with observed data and simulation output data (building energy model output) this study builds a Gaussian process for the given parameters. After building an emulator (Gaussian process), this study utilizes an emulator model to further the process. A Metropolis Hastings (MH) sampling method is then executed to sample from both the prior distribution and likelihood function. In MH sampling, one way to apply a model to data is to find the maximum a posteriori (MAP) parameter values. There are currently two updating schemes to perform MH sampling. This study used a component-wise updating scheme to derive the posterior distribution. For the likelihood function this study considered electricity.

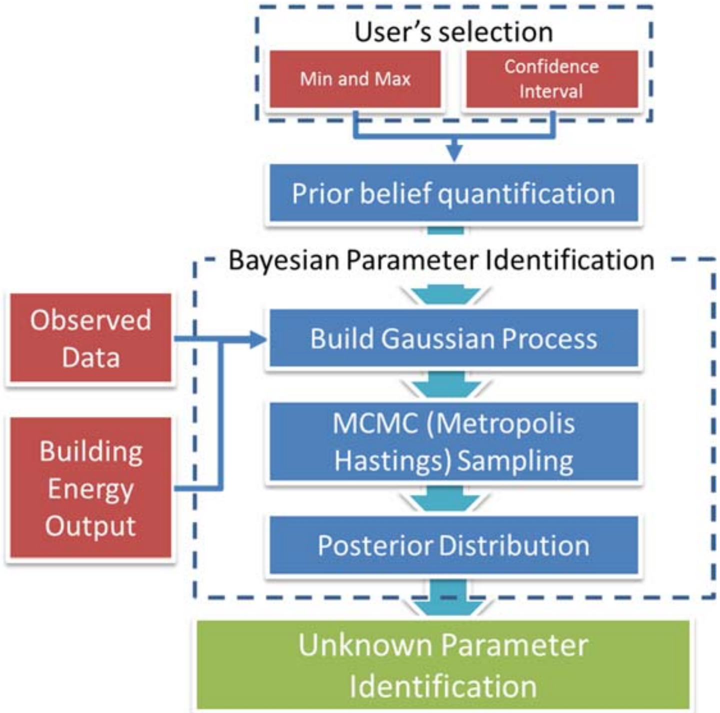


Figure 3-1. Proposed process for our study

3.1.1. Specification of prior distribution

Prior belief can be interpreted as a user's estimation or confidence level of a certain unknown parameter. In other words, it is an uncertainty of unknown parameters. As described before, some of the parameters in whole building simulation or building energy calculation are difficult to identify even with building expertise. To account for these uncertainties, it is necessary to consider how these subjective forms can be converted into objective forms that can be used mathematically. In probability, there are numerous distributions available (e.g., triangular, Poisson and Gaussian). Among these distributions, for the model of this study, normal distribution is assumed to represent prior belief over the unknown parameters. The reason why a normal distribution was selected is because of the characteristic of uncertainties that will be accommodated for each parameter. To mathematically accommodate the prior belief for each parameter, a normal distribution is the best fit for the occasion.

In probability, a normal or Gaussian distribution is a continuous probability distribution, defined by the following equation.

$$f(\theta_*) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(\theta_* - \mu)^2}{2\sigma^2}} \quad (3.6)$$

Where parameter μ is the mean and σ is its standard deviation of the Gaussian distribution function. In the energy model, μ can be a user's value selection of the unknown parameter σ can be the uncertainty associated with it. θ_* is the proposed value that we wish to determine the probability based on μ and σ . However, it is difficult to calculate both the mean and the standard deviation associated to uncertainties for each parameter in a building energy model. The issue is that standard deviation (σ) is an ambiguous value to be determined by the user. Therefore the term 'confidence level' can be used. The confidence level can be utilized and translated into a standard deviation value for the Gaussian distribution function

In statistics, there is a 68 – 95 – 99.7 rule or three sigma rule (Feller, 1971) or an empirical rule which states that for a normal distribution, 99.7% of the values lie within 6 standard deviations of the mean as described in Figure 3-2.

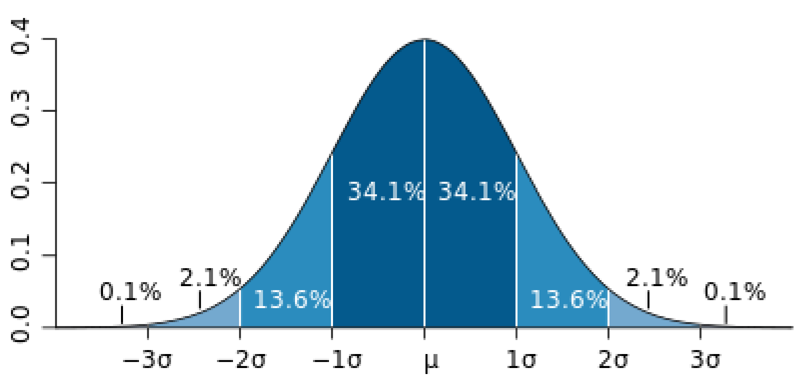


Figure 3-2. Standard deviation diagram Source: Wikipedia(Wikipedia, 2014)

In mathematical notation, these facts can be expressed as follows, where x is an observation from a normally distributed random variable, μ is the mean of the distribution, and σ is its standard deviation:

$$\begin{aligned} &\Pr(\mu - \sigma \le x \le \mu + \sigma) \approx 0.6827 \\ &\Pr(\mu - 2\sigma \le x \le \mu + 2\sigma) \approx 0.9545 \\ &\Pr(\mu - 3\sigma \le x \le \mu + 3\sigma) \approx 0.9973 \end{aligned} \tag{3.7}$$

For example, assuming we want to find the confidence interval with 2 sigma then the mathematical form will be as in the following equation.

$$\begin{aligned} \Pr(\mu - 2\sigma \le x \le \mu + 2\sigma) &= \Phi(2) - \Phi(-2) \\ &\approx 0.9772 - (1 - 0.9772) \approx 0.9545 \end{aligned} \tag{3.8}$$

Therefore 2σ will reside in a 95.45% confidence interval.

. To quantify a standard deviation based on confidence level and the user's selection of the parameter, this model requires the following input values for each unknown parameter: user's assumption of minimum and maximum value of the input and a confidence level (0%~100%), wherein 0% indicates that the user has absolutely no indication what the value might be and 100% indicates that the user has 100% confidence on the selection. With the confidence level we can translate this and utilize it in the z table (APPENDIX A).

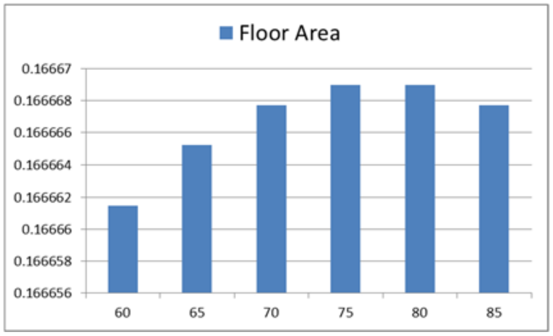
Moreover, to quantify the confidence level that was given by the user, the z table is used to determine the z score. After z score calculation, standard deviation σ will be calculated with number of sample n. Eq. (3.9) describes how z score is calculated based on number of sampling n, the mean of lower and upper limit of the parameter $\bar{\mu}$, and the mean value of minimum and maximum value of user's input μ. The following equation explains how the z score is calculated.

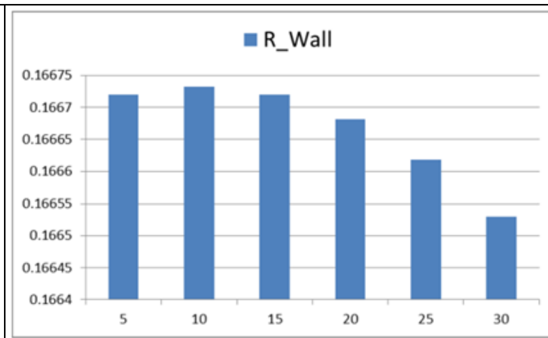
$$Z = \frac{\bar{\mu} - \mu}{\sigma(\theta_*)/\sqrt{n}} \tag{3.9}$$

Therefore deriving standard deviation as

$$\sigma(\theta_*) = \frac{\sqrt{n}(\bar{\mu} - \mu)}{Z} \tag{3.10}$$

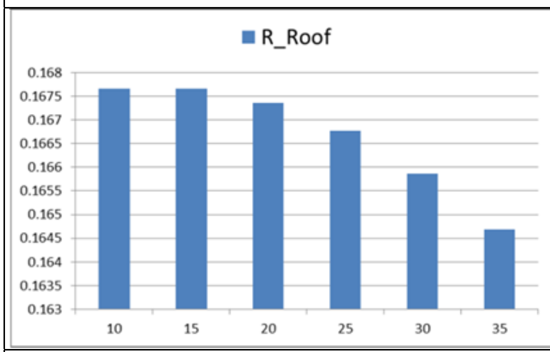
Figure 3-3 represents an example of prior calculation with a given user's input.

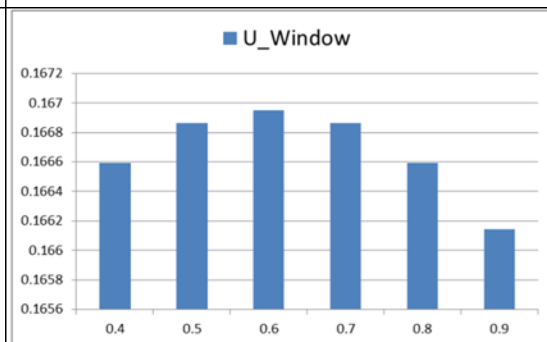




Min: 70, Max:85, CI: 10%

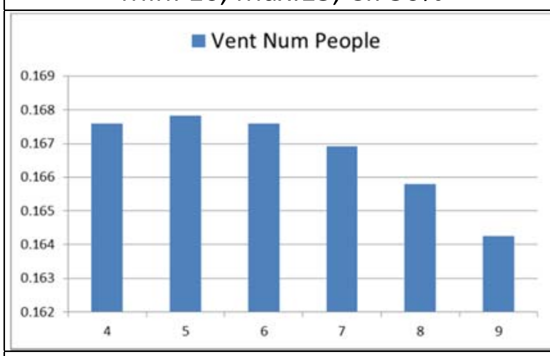
Min: 5, Max:15, CI: 30%

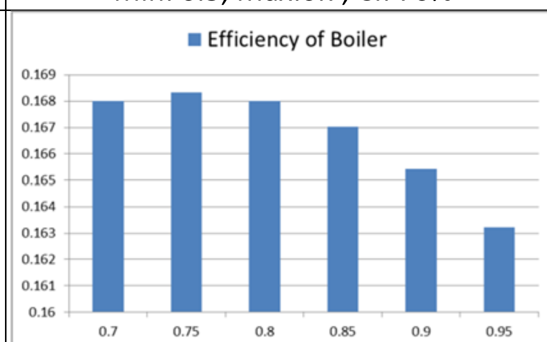




Min: 10, Max:15, CI: 50%

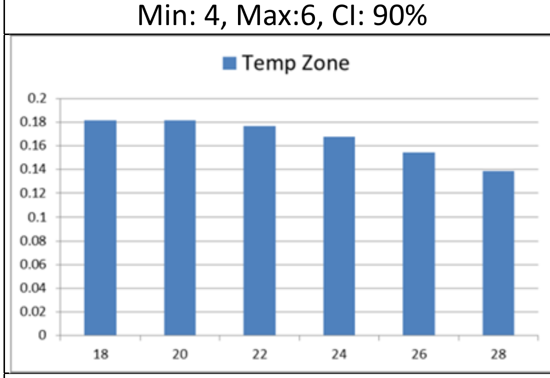
Min: 0.5, Max:0.7, CI: 70%





Min: 4, Max:6, CI: 90%

Min: 0.7, Max:0.8, CI: 95%



Min: 18, Max: 20, CI: 99%

Figure 3-3. Sample prior probability calculation with given input (minimum and maximum value of user's input and confidence interval)

3.1.2. Likelihood function

The likelihood function is a crucial element of the Bayesian-based parameter identification approach. Indeed, the likelihood function defines how likely the model or its emulator output from a given set of input parameters matches observation data. In this study, the likelihood function is defined by Eq. (3.11). Likelihood of the observation data y (e.g. utility data) given the outcome $\eta(\theta)$ of a set of model parameter value (x, θ) from emulator's output will be calculated using observation covariance matrix. Observation covariance matrix can be interpreted as a matrix that contains the noise that can be obtained while accumulating the observation data. Therefore, the likelihood is expressed by the probability of obtaining the observed outcome given the model parameter values.

$$\mathcal{L}(y|\eta(\theta)) \propto \exp\{-\frac{1}{2}\big(y-\eta(\theta)\big)^T \Sigma_y^{-1}\big(y-\eta(\theta)\big)\} \tag{3.11}$$

Where, $\mathrm{y} = \big(\mathrm{y}(x_1), \dots, y(x_n)\big)^T$ and $\eta(\theta) = \big(\eta(x_1, \theta), \dots, \eta(x_n, \theta)\big)^T$ and x represents the parameters that are known to users and no need to be identified (e.g. building geometry, weather data, etc.). Σ_y denotes the observation covariance matrix [25] and it influences the likelihood which will determine the acceptance of a plausible value for θ, therefore it is crucial to have appropriate observation noise (σ_y) when Σ_y is made of σ_y as $\Sigma_y = I_n * \sigma_\mathrm{y}$ where I_n is the identity matrix with n dimension (in our case, n equals 24 since there are electricity and gas data for each month of the year). One approach is to consider σ_y as unknown as θ and let the posterior distribution determine its value. Another approach is to select a value for σ_y using the results of a sensitivity analysis. One suggestion is to treat σ_y as one of calibration criteria that needs to be met, in terms of calibration.

3.1.3. *Posterior distribution*

In Bayesian statistics, the posterior probability of a random event or an uncertain proposition is the conditional probability that is assigned after the relevant evidence is taken into account. Similarly, the posterior probability distribution is the distribution of an unknown quantity, treated as a random variable, conditional on the evidence obtained from an experiment or survey. To derive a posterior distribution with given user input by using Bayesian approach, this study used the Markov Chain Monte Carlo method.

3.1.3.1. *Markov Chain Monte Carlo method (MCMC) (Steyvers, 2011)*

The application of probabilistic models to data often leads to inference problems that require the integration of a complex, high-dimensional distribution. MCMC is a general computational approach that replaces analytic integration by summation over samples generated from iterative algorithms. While MCMC is very much an active research area, there are now some standardized techniques that are widely used. Among those techniques, this study used a well-known technique, the Metropolis-Hastings algorithm. Supposing the goal is to sample from the target density $p(\theta)$, with $-\infty < \theta < \infty$. The Metropolis sampler creates a Markov chain that produces a sequence of values:

$$\theta^{(1)} \rightarrow \theta^{(1)} \rightarrow \cdots \rightarrow \theta^{(t)} \rightarrow \cdots \rightarrow \theta^{(n)} \tag{3.12}$$

Where $\theta^{(t)}$ represents the state of a Markov chain at iteration t and $\theta^{(n)}$ represents n^{th} sample where n is the number of sampling. After the sampling, the distribution of the sample reflects the target distribution $p(\theta)$. In the Metropolis procedure, we initialize the first state, $\theta^{(1)}$ to some initial value and then use a proposal distribution $q(\theta|\theta^{(t-1)})$ to generate a candidate point θ^* that is conditional on the previous state of the sampler. The next step is to either accept the proposal or reject it. Assuming the variable is univariate, the probability of accepting the proposal is:

$$\alpha = \min(1, \frac{p(\theta^{*})}{p(\theta^{(t-1)})}) \tag{3.13}$$

To make a decision on whether to actually accept or reject the proposal, we generate a uniform deviate u. If $u \leq \alpha$, we accept the proposal and the next state is set equal to the proposal: $\theta^{(t)} = \theta^{*}$. If $u > \alpha$, we reject the proposal, and the next state is set equal to the old state: $\theta^{(t)} = \theta^{(t-1)}$. We continue generating new proposals conditional on the current state of the sampler, and either accept or reject the proposals. This procedure continues until the sampler reaches convergence. Figure 3-4 illustrates the procedure for a sequence of two states.

However, for multivariate distributions involving N variables, we designed a N-dimensional proposal distribution and we either accept or reject the proposal as a block or as each components.

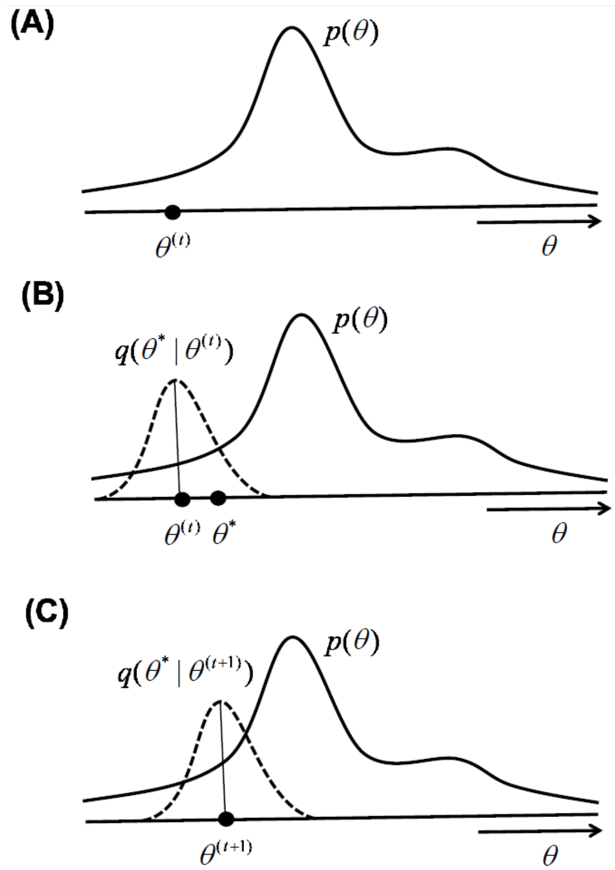


Figure 3-4. Illustration of the Metropolis sampler to sample from target density (Steyvers, 2011)

For a multivariate sampling situation, there are two ways to sample a posterior distribution. One method is called block-wise updating and the other is called component-wise updating. The Metroplis-Hasting (MH) sampler is a generalized version of the Metroplis sampler that can be applied with symmetric as well as asymmetric proposal distributions. (Steyvers, 2011)

3.1.3.1. Block-wise updating

This approach uses a proposal distribution that has the same dimensionality as the target distribution. For example, if the sampling distribution requires N variables, then this approach will generate a proposal set with N variables ($\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_N)$) and it either accepts or rejects the proposal as a block. The acceptance probability of a blockwise updating scheme is:

$$\alpha = \min(1, \frac{p(\boldsymbol{\theta}^*)}{p(\boldsymbol{\theta}^{(t-1)})} * \frac{q(\boldsymbol{\theta}^{(t-1)}|\boldsymbol{\theta}^*)}{q(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{(t-1)})}) \tag{3.14}$$

The below describes the process of this approach.

1. Set t = 1
2. Generate an initial value u and set $\boldsymbol{\theta}_t = \boldsymbol{u}(u_1, u_2, \dots, u_N)$
3. Repeat
    - $t = t + 1$
    - Generate a proposal $\boldsymbol{\theta}_*$ from previous probability (prior)
    - Evaluate acceptance probability α
    - Generate a u from a Uniform(0,1) distribution
    - If $u \leq \alpha$, accept the proposal and set $\boldsymbol{\theta}^{(t)} = \boldsymbol{\theta}^*$, else set $\boldsymbol{\theta}^{(t)} = \boldsymbol{\theta}^{(t-1)}$
4. Until t = T

3.1.3.2. Component-wise updating

A potential problem with the block-wise updating approach is that it might be difficult to find suitable high-dimensional proposal distributions. A related problem is that block-wise updating can be associated with "high rejection rates." Instead of accepting or rejecting a proposal for theta involving all of its components simultaneously, it might be computationally simpler to make proposals for individual components of $\boldsymbol{\theta}$, one at a time. This leads to a component-wise updating approach. For example, for bivariate distribution $\boldsymbol{\theta} = (\theta_1, \theta_2)$, the following describes the process of component-wise updating.

1. Set $t = 1$
2. Generate an initial value $\boldsymbol{u} = (u_1, u_2)$ and set $\boldsymbol{\theta}^{(t)} = \boldsymbol{u}$
3. Repeat
    - $t = t + 1$

- Generate a proposal θ_1^* from previous probability $q(\theta_1|\theta_1^{(t-1)})$
- Evaluate the acceptance probability $\alpha = \min(1, \frac{p(\theta_1^*,\theta_2^{(t-1)})}{p(\theta_1^{(t-1)},\theta_2^{(t-1)})} * \frac{q(\theta_1^{(t-1)}|\theta_1^*)}{q(\theta_1^*|\theta_1^{(t-1)})})$
- Generate a u from a Uniform(0,1) distribution
- If $u \leq \alpha$, accept the proposal and set $\theta_1^{(t)} = \theta_1^*$, else set $\theta_1^{(t)} = \theta_1^{(t-1)}$
- Generate a proposal θ_2^* from previous probability $q(\theta_2|\theta_2^{(t-1)})$
- Evaluate the acceptance probability $\alpha = \min(1, \frac{p(\theta_1^*,\theta_2^{(t-1)})}{p(\theta_1^{(t-1)},\theta_2^{(t-1)})} * \frac{q(\theta_2^{(t-1)}|\theta_2^*)}{q(\theta_2^*|\theta_2^{(t-1)})})$
- Generate a u from a Uniform(0,1) distribution
- If $u \leq \alpha$, accept the proposal and set $\theta_2^{(t)} = \theta_2^*$, else set $\theta_2^{(t)} = \theta_2^{(t-1)}$

4. Until $t = T$

3.2. Gaussian process

3.2.1. Definition

After Rumelhart's study (Rumelhart, et al., 1986) of utilizing neural networks on supervised learning, the empirical modeling of creating correlation among high-dimensional data gained attention among researchers. However, due to the uncertainty in neural network such as creating architecture, activation functions, deciding learning rate, etc., made many researchers to look for alternative solutions. One of the alternative methods that researchers found was by using the probabilistic framework. Neal (Neal, 1996) has shown that the Bayesian neural network, which the prior distribution over non-linear functions, can be categorized as a Gaussian process (MacKay, 1998). A Gaussian process is a continuous stochastic process which consists of random functions which is a set of random variables indexed by a continuous variable (Snelson, 2006). The Gaussian process assumes any set of function variables has a joint Gaussian distribution. Instead of using the parameterized networks in neural network, the networks are optimized and then replaced by simple matrix operations and therefore the Gaussian process allows to work directly without any decisions or assumptions needed as neural network. The Gaussian process is widely known for two properties. First, the Gaussian process can be determined by two major components: mean functions and covariance functions. These two components facilitates model fitting of entire data. Secondly, after determining the mean functions and covariance functions, it is easy to utilize the Gaussian process in a prediction or a regression problem. Popular ways to utilizing the Gaussian process can be used in either a regression problem or in a classification.

3.2.2. Why Gaussian process emulator?

To solve complex engineering problems, a simulation model can be utilized, and the simulation model can be interpreted as a model based on mechanistic and physical process drive models such as DOE2.2 or EnergyPlus. Normal whole building energy analysis can be time-consuming since simulation itself requires computing time to process numerous parameters. Hong (Hong, et al., 2008) compared several prototypes of buildings and showed DOE-2.1E v124's results along with EnergyPlus v2.1.0 and it took approximately 5.1 sec for DOE-2.1E and 158 seconds for an EnergyPlus simulation for a secondary school with a campus of 11 buildings and 79 thermal zones. However, a simulation model can be computationally expensive when dealing with a complex model. For computationally expensive simulation models, building an emulator can be another solution. In this case, an emulator is a statistical model that correlates the simulation model's input and output (Fricker, et al., 2010). Then, the emulator can be treated as a surrogate model or a meta-model for the actual simulation model. In model selection of the emulator, regression methods are often used to construct the emulator. Among various emulator models, Gaussian process methods appeared as a general choice. Sacks (Sacks, et al., 1989) first utilized the GP regression to model the simulation models' output and Currin (Currin, et al., 1991) utilized the Bayesian design/prediction methods. At the same time, utilizing neural network became one representative method to replace the Gaussian process.

Neural network (artificial neural network) is a computational model that was inspired by a biological nervous systems, such as the brain to be used to estimate or approximate functions which depends on inputs. Neural network was first proposed by McCulloch's study (McCulloch, et al., 1943) and then popularized by Rumelhart (Rumelhart, et al., 1986). Neural network can be a system composed of neurons (or nerve cells) assumed interconnected within the system. These neurons are known to compute values from inputs and are capable of machine learning and pattern recognition.

Several studies compared the Gaussian process and the neural network in various aspects. Mackay (Mackay, 1997) claimed that the Gaussian process is simply a smoothing device and in contrast, neural networks are intelligent models. The study suggested that many real-world data modeling problems can be well solved by sensible smoothing methods. Lilley (Lilley, et al., 2005) argued that the recurrent neural network can mimic the Gaussian process inference using only linear neurons, inhibitory recurrent connections, and one-shot Hebbian learning. Wang (Wang, et al., 2006) compared these two methods in forecasting and suggested that the artificial neural networks can be more useful when utilizing a greater size of data. However, this study also suggested that the Gaussian processes can be easily implemented to deal with non-stationarity. Table 3-1 compares the characteristic of the Gaussian process and the neural network approach. As shown in the table, computation burden can be higher for the Gaussian process. But it indicates higher the accuracy in the Gaussian process using a smaller set of training data. Also, the neural network can result in point estimate for the prediction of data while the Gaussian process can elicit the result of distribution which can be utilized in quantifying uncertainties in the result. Also, the training data can be limited due to accessibility of utility data of buildings. Therefore, in this case, the Gaussian process can be an appropriate method to utilize.

Table 3-1. Characteristic of the Gaussian process and the neural network

	Gaussian Process	Neural Network	Reference
Usability	Easy	Somewhat difficult	(Neal, 1996)
Scaling to a larger data sets	poor	good	
Utilizing kernels	yes	yes	
Accuracy	tractable and precise	depending on the training data	(Neal, 1996)
Computation	High	Reasonable	(Wang, et al., 2006)
Handling missing data	Good	Fair	(Wang, et al., 2006)
Estimation	Distribution estimates	Point estimates	

3.2.3. Methodology

A Gaussian process is a stochastic process for which any finite linear combination of random samples has a joint Gaussian distribution. It is important to note that it is a statistical model in which the property is inherited from a normal distribution.

In the previous section, this study has described the reason of choosing the Gaussian process as an emulator. As a result, the Gaussian process can be used as an emulator as a surrogate model for a whole building energy analysis, i.e., with DOE2.2 or EnergyPlus, since it can take less time to do a single run with higher accuracy. Regarding differences between emulation and simulation, emulation, in this case, is a process that mimics the behavior of a system, with no regard for how the system functions internally. Simulation, on the other hand, focuses on modeling of a system with consideration of the underlying state of the target system. Simulation can be referred to as a whole building energy analysis tool such as DOE2.2 or EnergyPlus.

This study's aim is to develop a Bayesian framework for unknown parameter identification of a building and therefore developing an emulator is an important aspect.

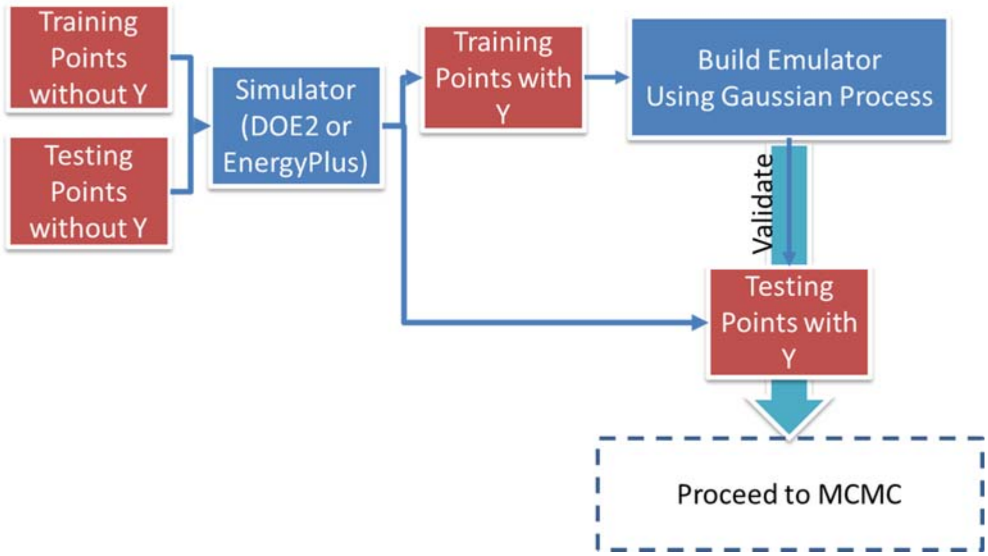


Figure 3-5 Flow process to build an emulator from simulator

Figure 3-5 represents the flow process for building an emulator of a whole building energy analysis tool. First, training points and testing points without y (in our case, energy consumption) are created by using a Latin hypercube sampling method. Latin hypercube sampling is a statistical way to generate a sample of a reasonable group of parameter values from a multidimensional distribution. This method can create samples that can cover the entire range of parameters so that it can be used as a plausible collection of parameters for an emulator. This will help an emulator to be more solid if it used for actual system evaluation. Once Latin hypercube sampling generates testing and training points, a simulator simulates its energy consumption and passes to the next process to build an emulator.

At this point, this study has built an emulator with a capability to provide Gaussian process regression. A Gaussian process can be divided into two sections, mean function and covariance function. If a Gaussian process is assumed to have a mean of zero, the covariance function completely defines its behavior. In this sense, the covariance matrix between all pair of training points is a key factor for building a Gaussian process regression (GPR). There are a number of covariance function kernels in reality. In our study, the squared exponential kernel was used to build the GPR in the following form.

$$\mathrm{Cov}\big((\mathrm{x},\mathrm{t}),(\mathrm{x}',\mathrm{t}')\big)=\frac{1}{\lambda_\eta}\exp\{-\sum_{k=1}^{p}\beta_k^\eta\,|x_{ik}-x_{ik}'|^2-\sum_{k\prime=1}^{\ell}\beta_{p+k\prime}^\eta\,|t_{ik\prime}-t_{ik\prime}'|^2\} \tag{3.15}$$

Where, λ_η stands for the reciprocal of the marginal variance of $\eta(\cdot,\cdot)$, β^η controls the dependence strength in each of the component directions of x and t. These are hyperparameters that control the covariance function. There are two ways of identifying these hyperparameters. One way is to treat these as same as θ and make an inference of these just like with θ. However, this process requires a considerable amount of time since it requires validation after each sampling method. Another method is to optimize them with conjugate gradient optimization (Rasmussen, et al., 2006). In this study, to avoid long computing times, an optimization method was used to identify all of the hyperparameters for a Gaussian process.

Figure 3-6 represents a more specific flow process of our approach with a Bayesian framework. As shown in the figure, the emulator used Metropolis Hastings Sampling for MCMC sampling and developed an emulator for its likelihood calculation. Normally, reference data that is utilized in the likelihood calculation is made of several outputs. For instance, if there is one reference data for each month in a year, that means there are 12 monthly data items that need to be specified as it is multi-dimensional data. For every dimension, it is known that a Gaussian process must be specified for each.

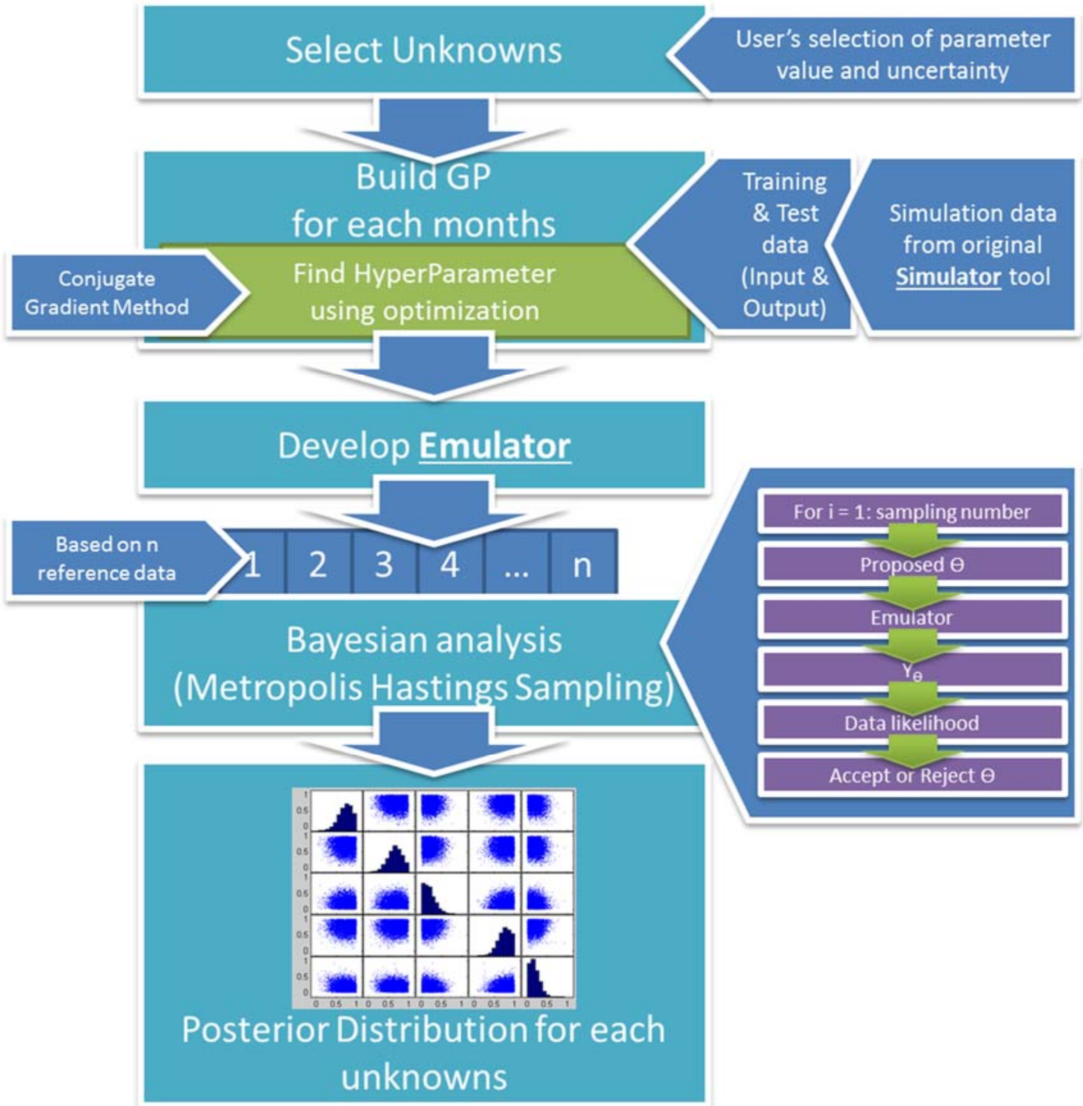


Figure 3-6. Specific flow of Bayesian framework

3.2.4. *Selecting the output*

The study presents the Gaussian process to be used as an emulator to replace whole building simulation and the Bayesian approach is used in the sampling methodology to output the posterior distribution. The difference between other optimization methods and the Bayesian approach for parameter identification challenges is that other optimization methods use a value for input and utilizes a point estimate for the output. On the other hand, the Bayesian approach inputs a prior distribution and outputs a posterior distribution and the value that has a highest occurrences within the posterior distribution can be selected as an answer for the parameter identification question. Figure 3-7 shows the different flow process of optimization methods and the Bayesian approach methods. The optimization methods are deterministic and therefore result in one set of identified parameter sets as an output. The Bayesian based approach results in posterior distribution and visualizes the possibility of parameters that can take place of unknown parameters. The result can be utilized not only for a parameter identification of a value but also can be utilized to quantify the uncertainty in the parameter identification by quantifying the posterior distribution. Therefore, it is more advantageous to utilize the Bayesian approach than the optimization methods.

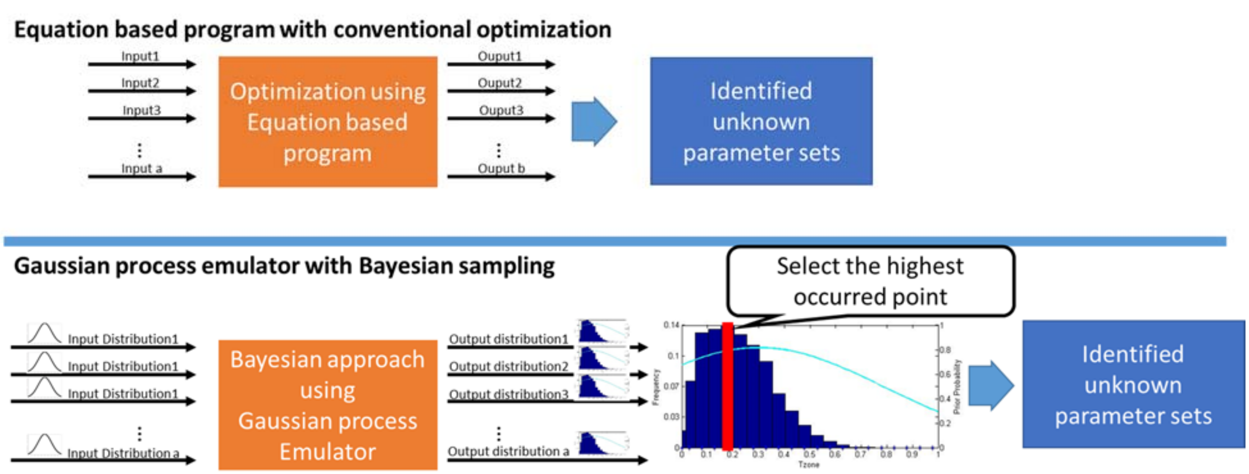


Figure 3-7. Difference between optimization methods and Bayesian approach

3.3. Preliminary case studies

For energy analysis, it is crucial to determine each parameter's impact on energy consumption. This is because its influence on energy consumption can determine which has a higher weight factor in the energy analysis and therefore can be used in the parameter identification based on this result. For instance, in a RC model, there are various input parameters needed within the model. However, to accommodate the reference data, tuning the most influential parameter alone will not be assumed as parameter identification. A good combination of all of the parameters will result in correct identification of the parameters. In order to achieve this goal, it is crucial to identify the sensitivity of each parameter. Table 3-1 represents sensitivity ranks among some of the variables for a RC model. Min, max value and increments indicate each parameter's option for selection. This in turn means this value represents the minimum and maximum value that each parameter can take. Further, the increments value indicates its gap between each selection (e.g., Zone Temp has 18, 20, 22, 24, 26, 28 options to select from).

Table 3-2. Sensitivity analysis for RC model parameters

	Min Value	Max Value	Increments	Energy consumption difference between options	Sensitivity Ranking
R Wall	5	30	5	7.70 % increase	3
R Roof	10	35	5	1.79 % increase	6
U Window	0.4	0.9	0.1	6.67 % increase	4
Number of People	4	9	1	8.43 % increase	2
Boiler Efficiency	0.7	0.95	0.05	6.14 % increase	5
Zone Temp (C)	18	28	2	15.40 % increase	1

Based on the sensitivity analysis, the temperature of zone has the highest impact on energy consumption. Based on this result, a Bayesian approach was performed to derive reference data. Figure 3-8 represents the posterior distribution result. Each bar represents occurrences from the posterior sampling process for the reference data. Along with a higher influence on the energy model, posterior occurrences have a higher result.

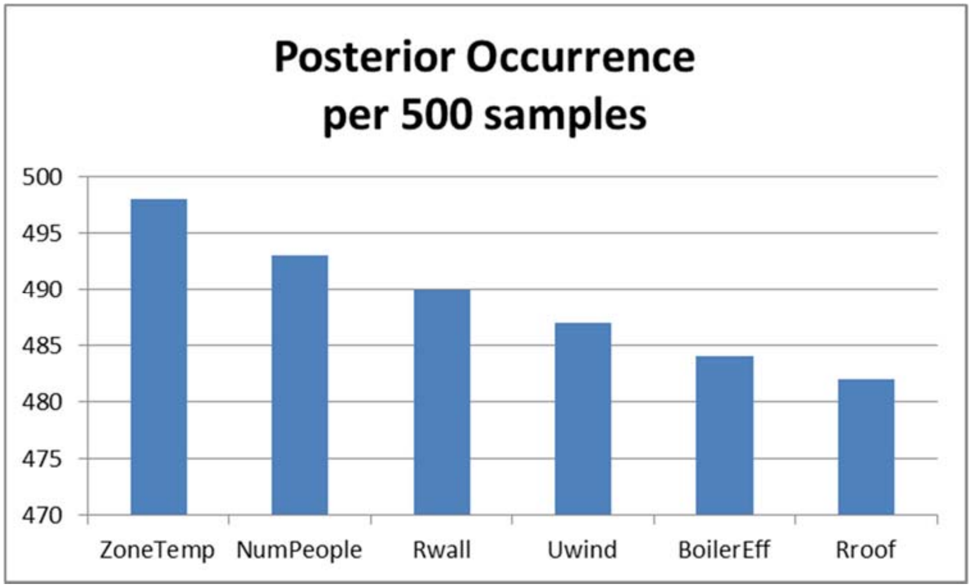


Figure 3-8. Posterior occurrence due to sensitivity of each parameter

3.3.1. Confidence level

Previously, this study described how to formulate prior probability for each parameter. In this section, we will validate this with actual results. Hypothetically, with a lower confidence level, it is more difficult to estimate an actual value. In contrast, with a higher confidence level it is assumed that the results will be more accurate. To test this theory this study performed case studies with two parameters in a RC model. This study selected the R value of roof and zone temperature for parameters to consider since they had the lowest and highest impact on the energy analysis. These two parameters were selected from among the other parameters to verify that the change of posterior distribution was due to the probability of the highest and lowest influential parameters. Figure 3-9 represents the result on the posterior distribution based on each confidence level. As shown in the figure, posterior occurrences tend to increase when there is an increase in confidence level. Additionally, the zone temperature tends to have higher posterior occurrences due to its higher impact on energy consumption. However, at a confidence level of 30%, the R value of a roof had more occurrences than one temperature. Since the Bayesian approach is based on random number generation within the algorithm; posterior occurrences can be lower with a higher impact parameter. Additionally, for the same reason, the posterior occurrence did not increase linearly along with an increase in confidence level.

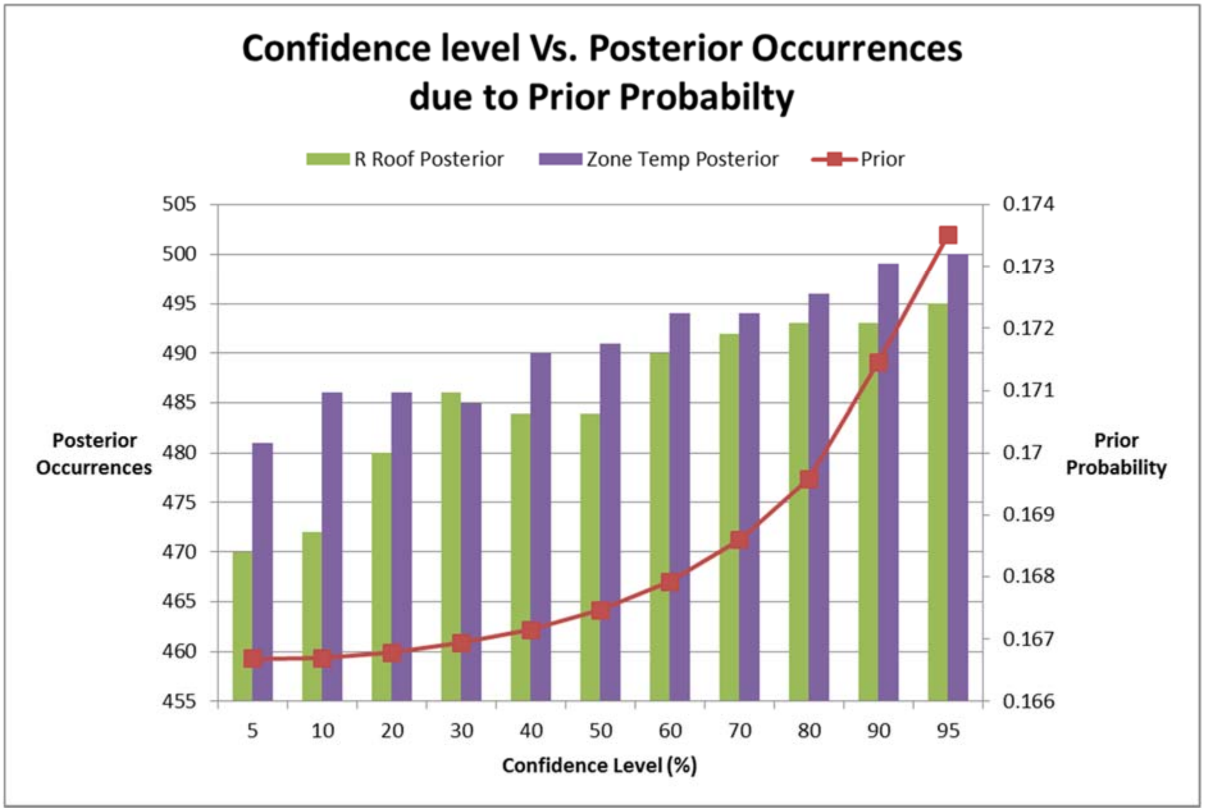


Figure 3-9. Posterior distribution Vs Confidence level

3.3.2. Different prior shapes

In the prior probability selection, there are many different prior probability distributions that can be applied to derive posterior distribution. Normal distribution can be the popular choice and, in Heo's study (Heo, 2011), prior probability distribution was defined as a triangular distribution which the probability peaks at the mean of the estimated value. In this study, different prior types of distribution were validated to verify the difference. Three shapes were tested; normal distributions with different confidence level, a uniform probability distribution and triangular probability distributions with different intensity. The case study was generated by using DOE2.2 model of large office building with two unknown parameters; lighting density and heating heat input ratio.

Figure 3-10 shows the result of posterior distribution using different prior shapes. On the top row, a normal distribution was used with a strong prior belief on the parameter's expected value (target value). The second row represents a posterior distribution result of a normal distribution with weaker confidence level. A uniform distribution on the prior belief was used in the third row. A triangular distribution with weak belief was used in the fourth row. And the last row displays the triangular distribution with strong belief on estimated parameter's value. The bar graph represents the posterior distribution result of each cases. As shown in the figure, nearly all of the posterior results identified the target value or found the nearest value despite of different prior shapes. However, the shape of the posterior distribution was slightly different. The priors with higher beliefs for both shapes exhibited more narrow posterior distribution around the target value. Other shapes showed more spread result. This can be predicted as priors with higher beliefs tends to accept more proposals around its target value while other shapes have less influence over proposal values. However, the difference between a triangular distribution and a normal distribution is that there is more flexibility in manipulating the prior shape. For a normal distribution, one can utilize the z-table using confidence level to derive the shape. But for a triangular shape, it is difficult to quantify the user's belief into the shape. Therefore, the normal distribution carries more advantage in quantifying the prior belief into the Bayesian methodology.

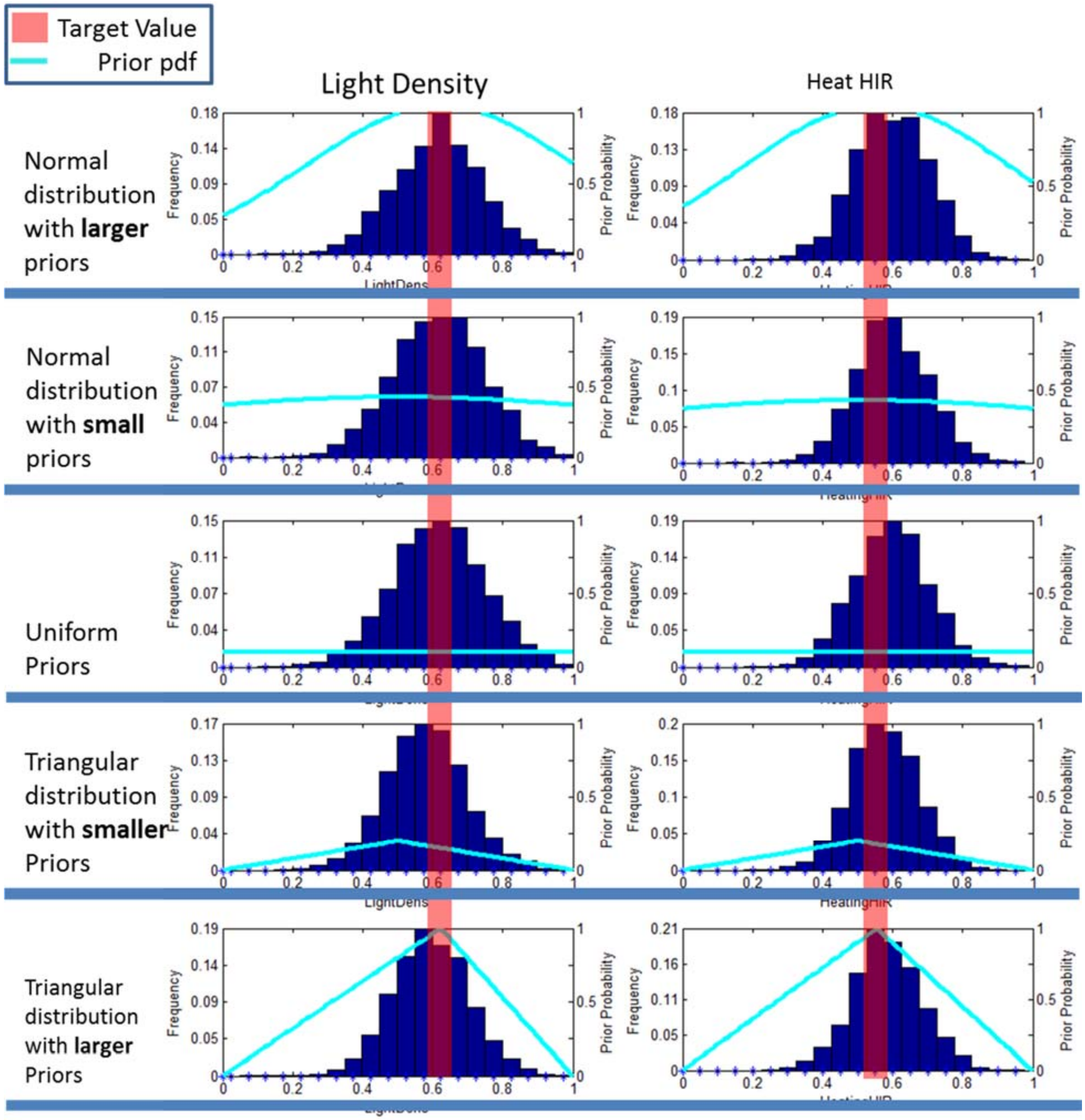


Figure 3-10. Posterior distribution of different prior shapes

3.3.3. Observation noise

In the Bayesian approach, the calculation of data likelihood is crucial to both accuracy and acceptance rate in deriving the posterior distribution. In Eq. (3.11), data likelihood mostly consisted of utilizing the reference data (utility data), the data from simulation, and observation covariance matrix which is consisted of observation noise. The observation noise represents an error or a noise that can be produced while obtaining the reference data. For instance, most of the buildings have a similar pattern or trend of energy usage in each year. However, energy usage in each year is slightly different. To accommodate this difference, Krarti (Krarti, 2010) used averaged yearly energy usages over 3 years or more into 1 year (12 months) data to be utilized in the energy analysis. Therefore, averaging the yearly data can generate an error due to the standardization of several years' worth of data.

To consider the error due to the observation of reference data, we can quantify the observation noise and use the observation noise within data-likelihood calculation. Pavlak (Pavlak, 2014) suggested that there are 2 ways to accommodate the observation noise. One method will be putting the observation noise as another unknown parameters within the Bayesian parameter identification and derive the posterior distribution along with unknown parameters. However, putting the observation noise as another unknown parameter will create more demand on computing because we have 1 more parameter to consider within the parameter identification. Also, the observation noise is sensitive to the overall accuracy and acceptance rate of the posterior distribution and therefore the uncertainty in the result of the parameter identification can increase. Another method is to pre-determine the value to be used in the Bayesian approach. We can generate several cases of the observation noises and select the appropriate value. In this study, we will evaluate various cases of the observation noise in the Bayesian approach and select the appropriate value for each case studies.

Three different building energy models were tested (DOE2.2 models of a small residential building, a medium office building, and a large office building). As the observation noise decreases exponentially, the acceptance rate decreases. However, after a certain value of the observation noise, the acceptance rate increases significantly untill it reaches 100%. This is because after some decrease in the observation noise value, the data-likelihood calculations fail to estimate the floating point numbers. Pavlak (Pavlak, 2014) mentioned that typical double precision computing environment is capable of calculating the floating point numbers on the order of 10^{-108} to 10^{308}. If the value is outside of this range, the calculation will cause the likelihood resulting posterior to be "Inf", "NaN", or "0". After failure of calculations, the algorithm rejects every proposed value and results in 0% acceptance. The proper selection of the observation noise can be simply done by selecting the observation noise value that will generate the acceptance rate of 25% ~ 45% as Gelman (Gelman, et al., 2013) suggested. If the value of the observation noise was set to accept larger range of proposed value then the posterior distribution will be dispersed and may not find the appropriate value of the unknown parameter. On the contrary, if the noise was set to accept a smaller range of value, it will create narrow posterior result. However, the posterior distribution may be distorted and there is a possibility of not exploring the entire range of unknown parameter values unless a larger sampling number was used. Additionally, if the observation noise was set to create an acceptance rate that was too high then the posterior distribution will only follow the behavior of the prior distribution since that is the only influence in the posterior probability calculation. On the other hand, if the observation noise is set to generate an acceptance rate that is too small, the posterior distribution will only mimic the behavior of the difference between reference data and simulated or emulated data without any impact from prior belief. So, it is important to set the observation noise that will balance the prior belief and the data likelihood. Therefore, this study has selected the observation noise based on the sensitivity analyses which will derive the acceptance rate between 25% ~ 45%.

Figure 3-11 shows the sensitivity analysis results of the acceptance rate for different observation noise cases. The x-axis provides the level of the observation noise and left y-axis lists the acceptance rate matching each observation noise case. And for the case studies, this study utilizes the 25~45% acceptance rate threshold to select the observation noise for each of the building cases. For DOE2.2 energy models, the appropriate observation noise is found to be in the range of $e^{-8} \sim e^{-11}$. As shown in Figure 3-11, when the acceptance rate decreases linearly, the observation noise decreases exponentially. And after the observation noise of $\sigma_y = e^{-16} \sim e^{-20}$, the acceptance rate produces 0% acceptance rate and this is due to the failure of calculation as Pavlak (Pavlak, 2014) suggested.

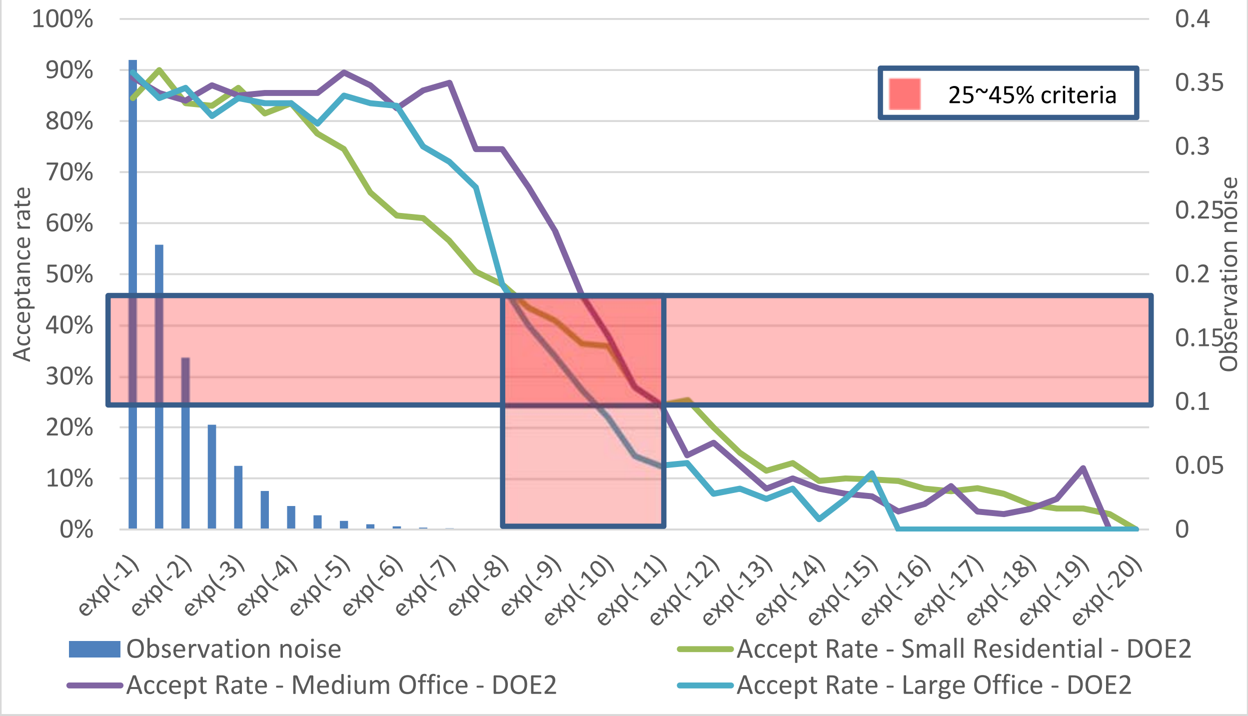


Figure 3-11. Sensitivity analysis of the observation noise on various DOE2.2 model cases

Figure 3-12 illustrates the acceptance rate results for different observation noise cases of

DOE2.2 small residential building. For comparison, the prior belief was selected randomly with 70% confidence level. As shown in Figure 3-12, the observation noise increases with the acceptance rate. For higher acceptance rates, the posterior results mimic the behavior of prior shapes. This is due to the prior belief carries much stronger influence than the likelihood in the posterior probability calculations. Therefore, the overall results of posterior distribution follow the shape of the prior belief. When the observation noise decreases exponentially, the likelihood has more influence than prior. The posterior is only impacted by the likelihood. This causes the narrow results of the posterior distribution. Highlighted area within the Figure 3-12 suggests the observation noise that is within 25 ~ 45% acceptance rate.

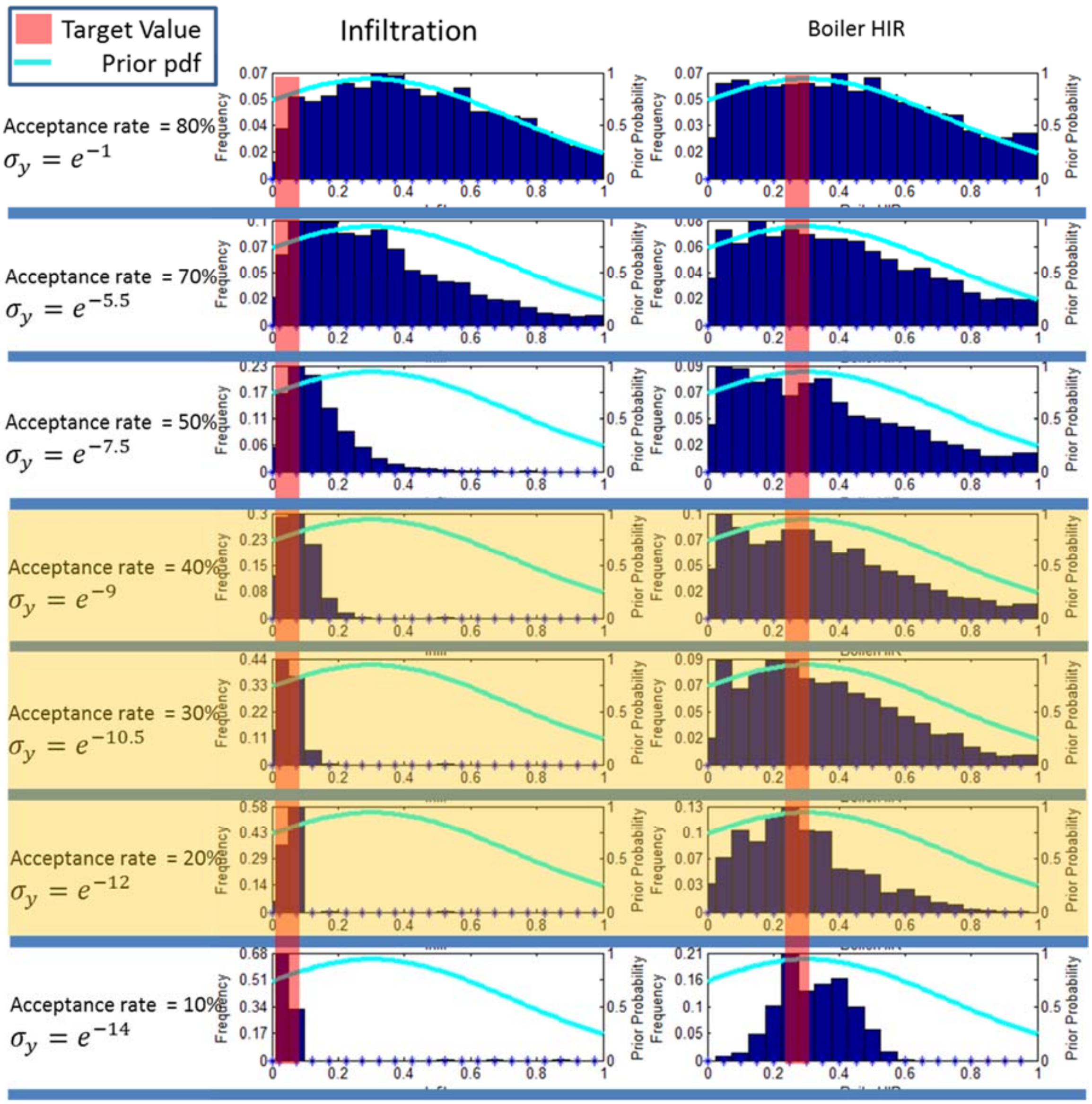


Figure 3-12. Acceptance rate result at different observation noise for DOE2.2 small residential case

CHAPTER 4: Validation of Bayesian framework

4.1. Introduction

Due to the evolution of building energy analysis, building energy models are now closer to true actual thermodynamic behavior. Results have been numerically proven to reflect actual buildings. In building energy modeling especially, building energy analysis can closely represent actual buildings. However, as building energy modeling has become more sophisticated, the modeling requires more inputs to more precisely reflect actual buildings. This can be interpreted as adding uncertainty thru these inputs. Building energy modeling has become increasingly important for assessing the current situation and for future plans, in not only the design phase of building construction, but also during phases for commissioning, maintenance, and all other phases. Uncertainty in such energy modeling can thus lower the accuracy of increasingly utilized analysis. Therefore determining uncertainties in building energy models is critical.

4.2. Literature review

Eisenhower (Eisenhower, et al., 2011) studied uncertainty and sensitivity decomposition in building inputs for energy models. The study configured the output distribution of office electricity consumption at peak load for 5,000 parameter realizations. They found that the two peaks were represented as an output distribution as shown in Figure 4-1. This can be interpreted as mis-representation of these parameter realizations because at a certain range, 5,000 parameter realizations will not represent its outcomes. The study claims that these mis-representations were derived from not accounting for uncertainties and sensitivities in the building input parameters.

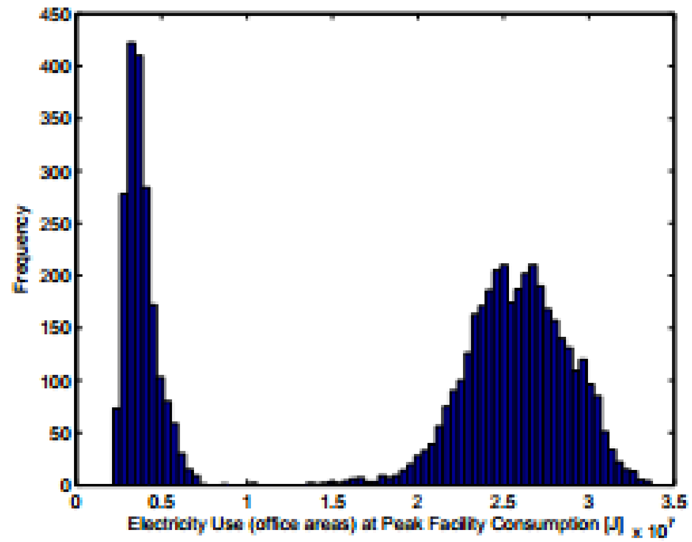


Figure 4-1. Output distribution of office electricity consumption at peak load for 5000 parameter realization Source: Eisenhower, et al (Eisenhower, et al., 2011)

Eisenhower used building number 7230 (the Atlantic Fleet Drill Hall) at the Naval Station Great Lakes, Great Lakes, Illinois for their study. The study chose almost all numeric parameters in the EnergyPlus Input file except for a few parameters that needed to be constant such as size, shape, orientation, etc. The resulting number of input parameters of 1,009 varied ± 10 or 20% of their nominal value and resulted in creation of 5000 in realization. With these realizations, a standard input-output sensitivity analysis was performed and sensitivity decomposition for facility energy was done as shown in Figure 4-2 (figure on the right). The figure on the left describes the example distributions for cooling (top) and heating (bottom) when varying the input parameters by either 10% or 20%. This type of analysis, including the decomposition, is valuable for identifying which subcomponents of a model need more attention during building design or model calibration.

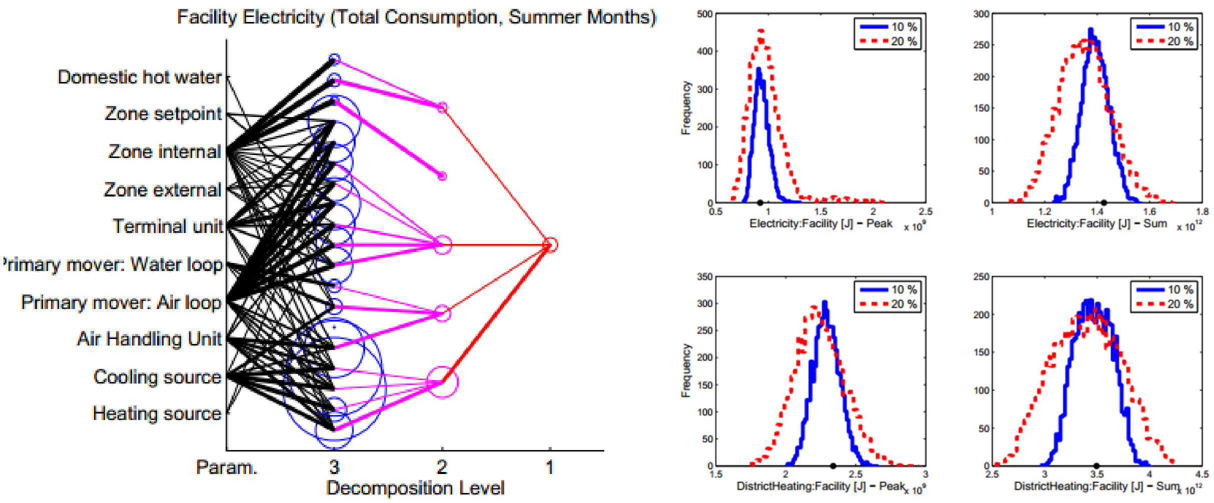


Figure 4-2. Sensitivity decomposition and its result distribution for facility energy for input parameters of 1009 from EnergyPlus Source: Eisenhower, et al (Eisenhower, et al., 2011)

Eisenhower (Eisenhower, et al., 2012) then studied a model-based optimization of building design and operation. Figure 4-3 represents its process flow of proposed optimization of building energy models. Based on Eisenhower's (Eisenhower, et al., 2011) study, this study performed uncertainty and sensitivity analysis in the beginning phase and identified uncertain parameters in the model. A reduced form of a meta-model was created by omitting the influence of a chosen set of parameters. Optimization was then performed on many different meta-models of the original baseline energy model that began with 1009 optimization parameters based on the thermal comfort (PMV) and the annual energy consumption for the facility.

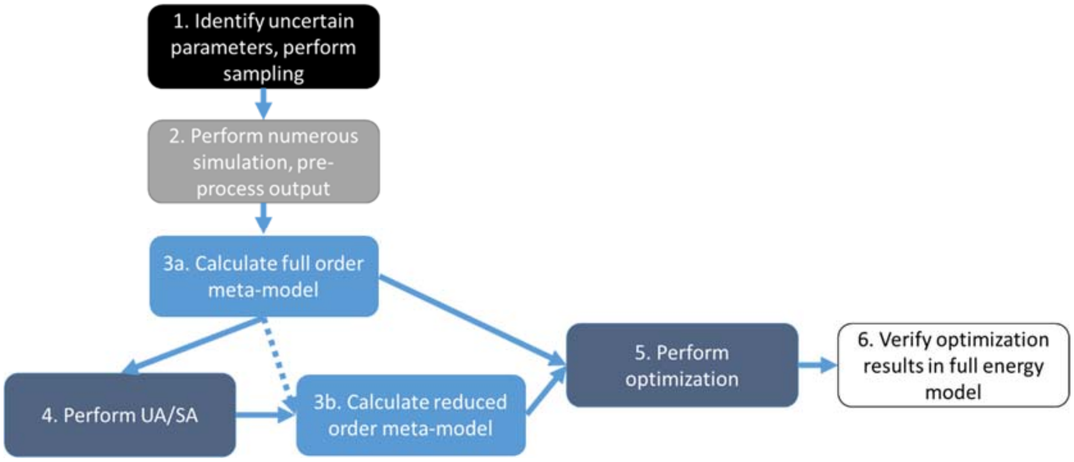


Figure 4-3. Process flow of suggested optimization of building energy models (Eisenhower, et al., 2012)

Two optimization methods were applied. This study utilized a gradient-based optimizer (e.g., an interior point (IP) method) that is a Primal-Dual Interior Point algorithm with a filter line-search method for nonlinear programming and derivative-free (DF) optimization that contains a Mesh Adaptive Direct Search (MADS) algorithm. These optimizations were done on a meta-model created to define the unknown parameters. In order to build a meta-model, Support Vector Regression (SVR) was selected from among various approaches (e.g., artificial neural networks, genetic programming, and Bayesian Networks). The figure below represents the data from EnergyPlus and SVR meta-model for 5,000 different building design and operation scenario iterations within 20% of the baseline case.

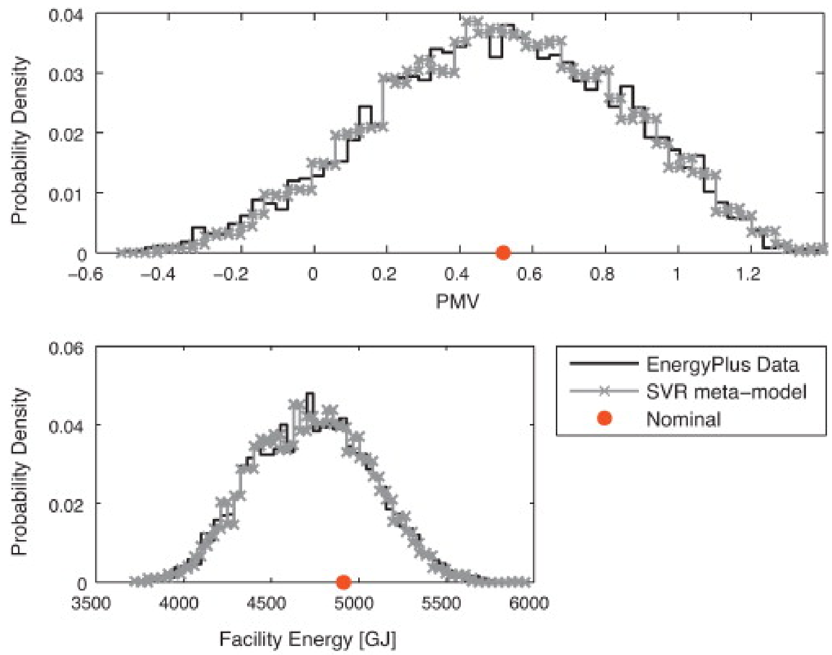


Figure 4-4. Output from EnergyPlus and SVR meta-model for 5000 different building design and operation scenario within 20% of the baseline case Source: Eisenhower, et al (Eisenhower, et al., 2012)

Table 4-1. Number of function evaluations for the various optimization experiments using both the IP (IPOPT) and the DF (NOMAD) methods. Eisenhower's study (Eisenhower, et al., 2012)

Optimization experiment	IPOPT	NOMAD
Full model [1009,C1]	6048	113,766
Full model [1009, C2]	23,184	1,000,000
Full model [1009, C3]	42,336	1,000,000
Top 20 [20, C3]	380	4944
Influence both PMV and energy [6, C3]	36	436
Schedule parameters [180, C3]	10,203	1,000,000
Material properties [142, C3]	35,814	854,212
Outdoor air controller [16, C3]	345	2130
Variable volume fan [48, C3]	2256	89,403
Top 7 [7, C3]	84	1312
Top 7 [7, C3] E+ in the Loop	NA	726

Table 4-2. The twenty most influential parameters on both comfort and energy consumption in Eisenhower's study ((Eisenhower, et al., 2012)

Reduced parameter number	Label	Parameter description
1	B	Minimum outside air fraction in occupied hours
2	B,C	AHU 1/2 winter (1/1 to 4/15) supply air temp. setpoint
3	B,C	AHU 1/2 summer (4/16 to 8/15) supply air temp. setpoint
4	B,C	AHU 1/2 winter (8/16 to 12/31) supply air temp. setpoint
5	C	Hot water supply temperature setpoint
6	C	Weekday zone temp. setpoint from 12:00 am to 6:00 am
7	B	People activity level (in W) in office area
8	B	People activity level (in W0 in Drill Deck
9		AHU4 summer (4/16 to 8/15) supply air temp. setpoint
10		Domestic hot water supply temp. setpoint
11	B	Water equipment target temp. setpoint
12		Domestic hot water usage fraction from 11:00 to 12:00
13		Domestic hot water usage fraction from 12:00 to 13:00
14		Domestic hot water usage fraction from 13:00 to 14:00
15		Domestic hot water usage fraction from 16:00 to 17:00
16		AHU2 return fan maximum flow rate
17		AHU1 minimum outside flowr rate
18		AHU2 minimum outside flow rate
19		AHU3 minimum outside flow rate
20		Chiller reference COP (coefficient of performance)

Table 4-1, the right table represents the number of function evaluations for the various optimization experiments using both the IP and DF methods. The left table represents the 20 most influential parameters for both comfort and energy consumption. Table 4-2 below represent optimization results using eight meta-models as described in the above table. The dotted line in Figure 4-5 represents the baseline of the EnergyPlus model and the bar graph represents the optimization results for these eight meta-models using two different optimization methods.

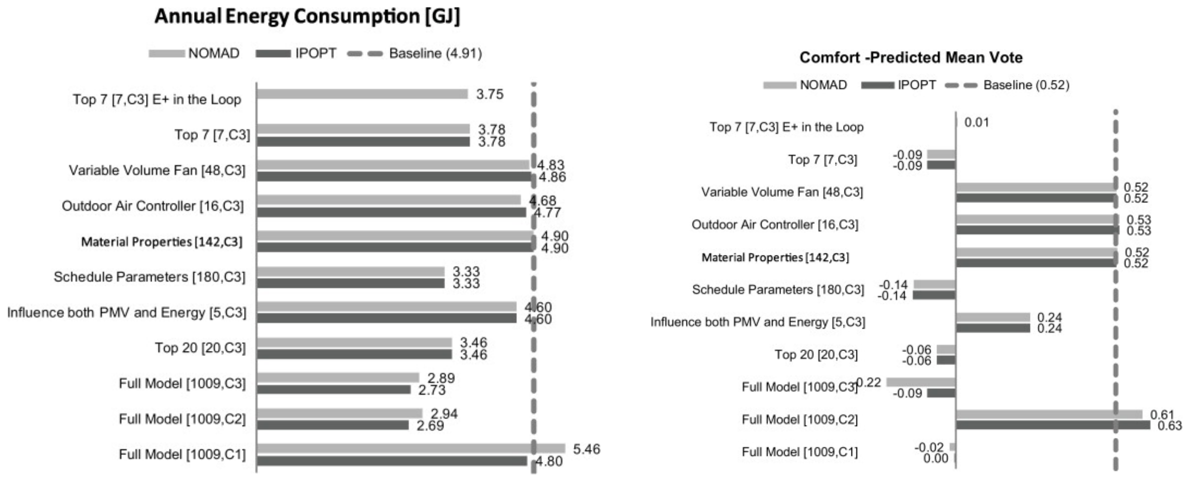


Figure 4-5. Optimization result for these 8 meta-models using 2 different optimization methods Source: Eisenhower, et al (Eisenhower, et al., 2012)

Pavlak (Pavlak, 2014) developed reduced order models such as ISO 13790 (Simple Hourly Method) and inverse gray-box thermal RC networks to exploit nonlinear least squares and Bayesian parameter estimation. His study was aimed at bringing a model-based methodology into a supervisory model predictive controller to optimize building operation in consideration of energy prices, demand charges, and ancillary service revenue. Considering these items the study was able to bring about a total savings of up to thirteen percentage points. When comparing Bayesian and nonlinear least squares parameter estimation, it was determined that the Bayesian approach had more advantages for various reasons. While a nonlinear least squares method finds a point estimate for the parameter, a Bayesian approach had more advantages for estimating unknown parameters. The data likelihood calculation in the Bayesian approach of the study included a sensitivity analysis on measurement error (σ_{ϵ}). The figure below represents the results. As measurement error decreased from 150 kW, the posterior maximum increased since the probabilities were concentrated among fewer parameter combinations.

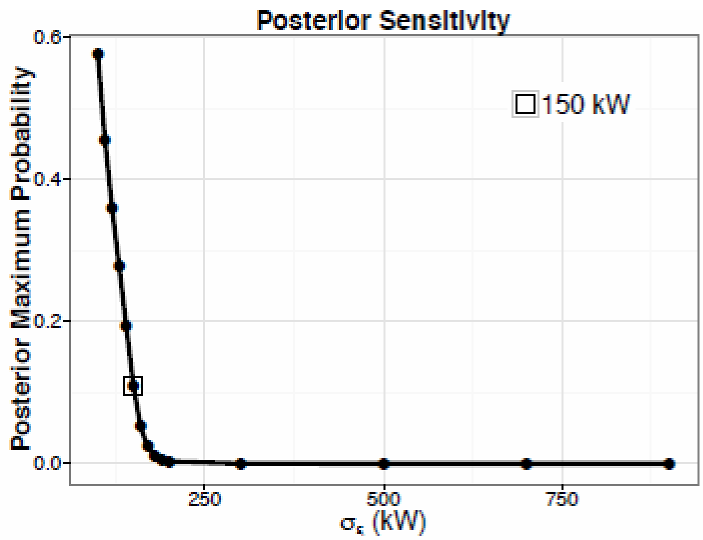


Figure 4-6. Sensitivity result on measurement error (σ_ϵ) Source: Pavlak (Pavlak, 2014)

Table 4-3. Performance summary metrics result (Pavlak, 2014)

		RMSE	**MBE**	**Cum. % Err.**
	NLSQ	10,250 W	347 W	0.869%
Load	Bayesian (σ_ϵ assumed)	10,692 W	1,802 W	4.491%
	Bayesian (σ_ϵ estimated)	10,244 W	-24 W	-0.059%
	NLSQ	0.2974 K	0.0360 K	0.191%
Temp.	Bayesian (σ_ϵ assumed)	0.2999 K	0.0367 K	0.195%
	Bayesian (σ_ϵ estimated)	0.2991 K	0.0386 K	0.205%

Pavlak also performed estimation on measurement error as a sixth parameter to determine an appropriate value. Table 4-3 shows the results of a comparison between a nonlinear least square method and a Bayesian (with measurement error assumed and estimated) approach. During this analysis, it has been found that Bayesian methods required 100 times more CPU time than the traditional parameter estimation and concluded that overall the study would not benefit from Bayesian methods. In an online operational setting, however, Bayesian methods maybe preferred if uncertainty quantification is desired or high levels of sensor noise are expected. Additionally, Bayesian methods may also have additional benefit in more complex, higher-dimensional models with multiple sources of noise and uncertainty.

As this study suggests, a traditional Bayesian approach has a drawback of taking too much CPU time to process. This can be caused by various reasons. One of the most significant comes from estimating parameters such as hyperparameters, measurement errors, etc. Hyperparameters are free parameters that allow for flexible customization of the Gaussian process in its covariance function. A traditional Bayesian approach requires these parameters to be estimated along with other parameters. However, Rasmussen (Rasmussen, et al., 2006) claimed that hyperparameters can be optimized by using a gradient-based optimization algorithm.

Decreasing CPU time while maintaining accuracy should be considered as a key factor in the Bayesian approach. In this study, the proposed Bayesian approach is suggested as shown in Figure 4-7. As shown in the figure, traditional Bayesian parameter identification utilizes the Gaussian process and sampling process at the same time and creates a bottleneck procedure that results in a high computational demand. The bottleneck arises from building the Gaussian process and sampling method in the same procedure. This results in numerous number threads creating the Gaussian process along with numerous sampling numbers. When a new proposed value for a hyperparameter is given, this elicits the creation of a new Gaussian process, which is re-evaluated every time. In contrast, the proposed Bayesian approach will eliminate this bottleneck process by building the Gaussian process as preparation for the sampling process. Therefore, the sampling process will use the emulator in its data-likelihood evaluations. This will eliminate the need to re-evaluate hyperparameters and will result in less CPU time.

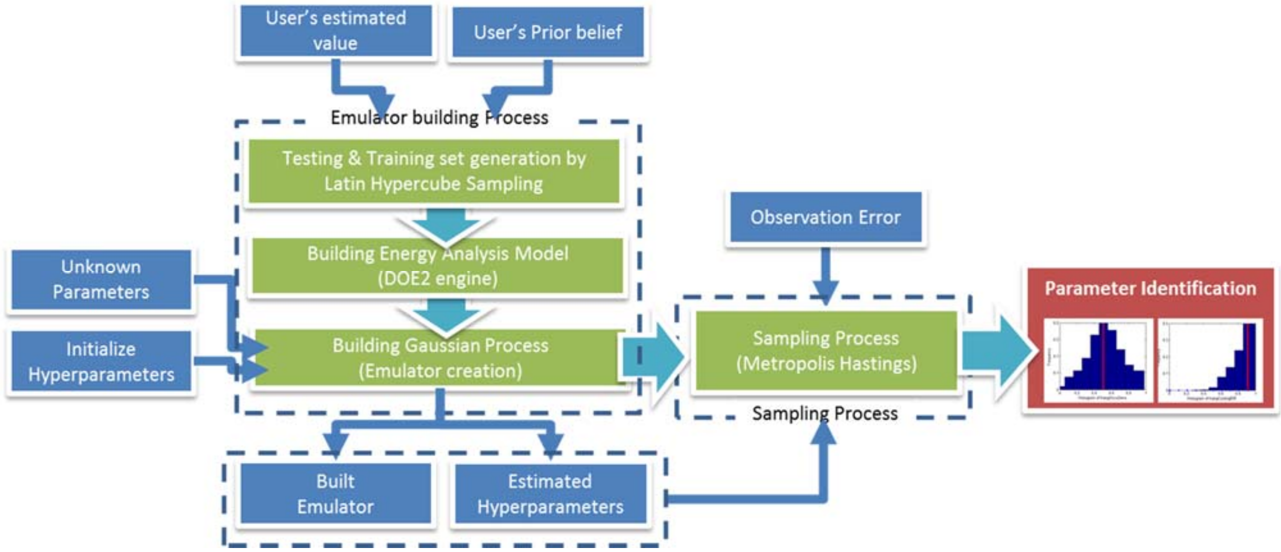


Figure 4-7. Proposed Bayesian parameter identification approach

4.3. Validation of RC model

4.3.1. Description of RC model

To validate this approach prior to exploring it in the whole building energy simulation, this study proposes to be tested in the simpler energy model. This simple model requires being efficient in computational time and accurate results. To achieve this goal, this study used a 3R2C (three resistance and two capacitance) model to calculate zone temperature (T_{zone}) by using hourly weather data (Park, 2012). This study calculated energy consumption of the zone by associating T_{zone} and thermal load equations. Figure 4-8 represents the overall process of a RC model.

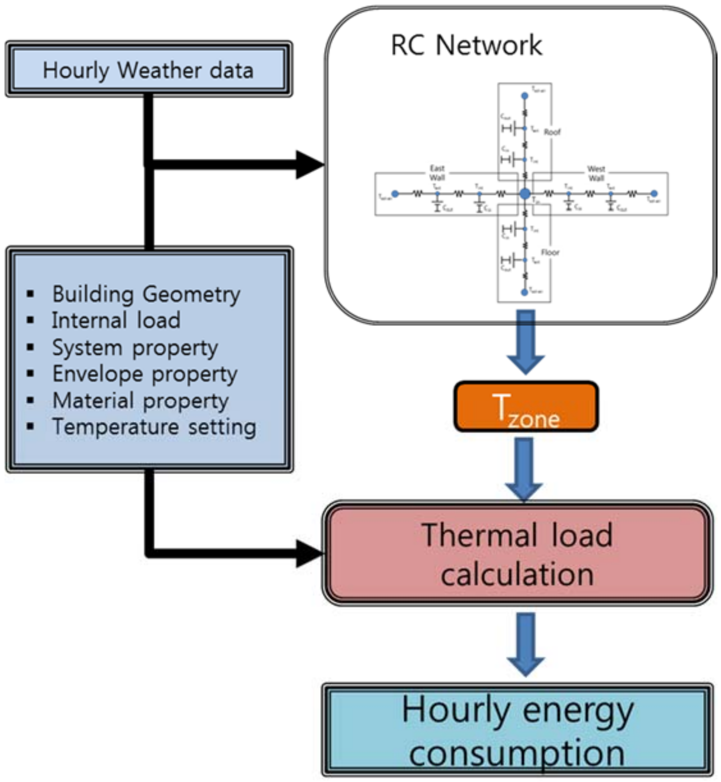


Figure 4-8. Flow chart of process of RC model

4.3.1.1. RC network model

Figure 4-9 illustrates the schematic of the RC thermal network for this study. This RC network calculates the convective heat transfer exchanges between surfaces including walls, a floor and a ceiling, which are accounted for by the RC model and are combined with indoor air heat balance. To account for the incident solar radiation impact on exterior surfaces, the sol-air temperature is assumed to be the outdoor air temperature.

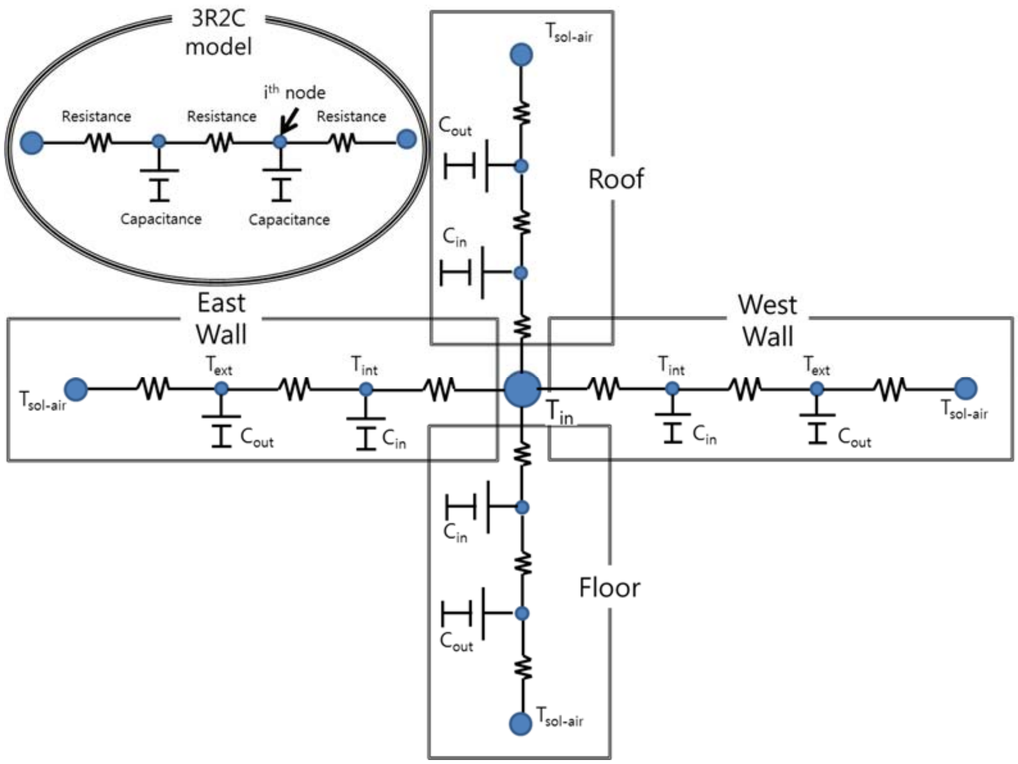


Figure 4-9. Schematic of RC energy model

$$A * \frac{\rho C_p}{2} * \frac{T_i^{t+1} - T_i^t}{\Delta t} = kA * \frac{T_{i-1}^{t+1} - T_i^{t+1}}{\Delta x} + hA * (T_i^{t+1} - T_{in}^{t+1}) \tag{4.1}$$

Where,

T_i: temperature at i^th node [C], T_{in}: indoor temperature [C]

Δt: timestep, t: time

k: thermal conductivity for each component [W/m-C]

h: convective coefficient for each component

A: surface area of each component [m^2]

The total convective heat transfer equation for the zone is described in the equation Eq. (4.2) which allows us to calculate T_{zone}. For more information, please see (Park, 2012).

$$\dot{m}c_p \frac{T_{zone}^{t+1} - T_{zone}^{t}}{\Delta t} = h_{EW} * A_{EW}(T_{EW}^{t+1} - T_{zone}^{t+1}) + h_{WW} * A_{WW}(T_{WW}^{t+1} - T_{zone}^{t+1}) + h_{roof} * A_{roof}\left(T_{roof}^{t+1} - T_{zone}^{t+1}\right) + h_{floor} * A_{floor}\left(T_{floor}^{t+1} - T_{zone}^{t+1}\right) \quad (4.2)$$

Where,

$\dot{m}$: air mass [kg]

c_p: specific heat of air [J/kg-K]

h_{EW}, $h_{WW}, h_{roof}, h_{floor}$: Convective coefficient for each component

$A_{EW}, A_{WW}, A_{roof}, A_{floor}$: Area for each component [m^2]

$T_{EW}, T_{WW}, T_{roof}, T_{floor}$: Temperature for each component [C]

Δt: timestep, t: time

4.3.1.2. *Energy consumption calculation for RC network model*

With calculated T_{zone}, we can calculate the heat loss from building envelope.

$$Q_{cond} = A * U * (T_{zone} - T_{out}) \quad (4.3)$$

Where,

Q_{cond}= Total hourly rate of heat loss through wall, roof, glass, etc [Btu/hr]

U= Overall heat-transfer coefficient of walls, roof, glass, floor, etc [Btu/hr ft^2 F]

A= Net area of walls, roof, ceiling, floor, or glass [ft^2]

T_{zone}= Temperature of zone [F]

T_{out}= Temperature of outside [F]

Heat loss can occur due to infiltration and ventilation. Ventilation rate is based on occupancy method and on the volume flow rate of infiltrated air, CFM, which determines the total heat loss from infiltration, which can be calculated as

$$V = N * Vent_{rate} \tag{4.4}$$

$$Q_{vent} = V * \rho_{air} * c_p * (T_{zone} - T_{out}) * 60$$

Where,

V = Ventilation air in [CFM]

N = Number of people in space

$Vent_{rate}$ = Ventilation rate [CFM/person]

Q_{vent} = Total hourly rate of heat loss through infiltration and ventilation [Btu/hr]

ρ_{air} = Density of air [lbm/ft^3]

c_p = Specific heat of air [Btu/lbm-F]

Additionally, internal heat gain can be calculated.

$$Q_{int} = 140 * N_{people} + LPD * A + EPD * A \tag{4.5}$$

Where,

Q_{int} = Total hourly rate of internal heat gain [Btu/hr]

N_{people} = Number of people

LPD = Lighting power density [W/ft^2]

EPD = Equipment power density [W/ft^2]

A = Area of the space [ft^2]

Once the components of heat loss and heat gain are calculated, total heating load on that zone can be calculated by summing up the heat loss from conduction and ventilation and internal gain. Total heating energy consumption can be obtained by dividing the total load by system efficiency.

$$Q_{total} = Q_{cond} + Q_{conv} - Q_{int} \tag{4.6}$$

$$HeatingE = \frac{Q_{total}}{System_{eff}}$$

Where,

Q_{total}: Total heating load [Btu/hr]

Q_{cond}: Conduction heat loss [Btu/hr]

Q_{conv}: Convection heat loss [Btu/hr]

Q_{int}: Internal heat gain [Btu/hr]

$System_{eff}$ = Efficiency of the system

$HeatingE$ = Heating energy consumption [Btu/hr]

4.3.2. Sensitivity analysis of parameters

To validate the methodology that this study proposes, the RC model for a building was created and tested in a closed-loop situation. By the term "closed loop situation," we mean that we are testing our model with a built environment for the experiment. To achieve simplicity and accuracy for model validation, instead of using a real data from an actual environment, we use simulated data from a built model (such as the RC model) as a reference data. This can be achieved by using a given RC model to obtain a reference data by using a preset value of parameters. After obtaining the reference data, we can set some of the parameters as unknown and try to identify those parameters with our proposed methodology.

In order to achieve this, sensitivity analysis of the parameters for the RC model has to be performed in advance. This is needed because the influence of parameters on energy consumption is different for each different parameter. Also, parameters that have a negligible impact on energy consumption can be ignored in the tuning process for the energy modeling. This is because a normal intuitive calibration focuses on parameters with higher impact on energy consumption rather than focusing on parameters with the least impact.

Therefore, this study performed sensitivity analysis on all of the parameters that are needed in order to run a RC model. The below table presents the results. The results showed that the window-to-wall ratio (WWR) had the most influence followed by zone temperature. In contrast, coefficients related to convection and solar had very little impact on energy consumption. With these results, this study performed unknown parameter identification in later sections.

Table 4-4. Sensitivity analysis of input parameters for RC model

Variable Name	Description	Rank	Impact on Energy Consumption (kWh) per 10% variance
'FloorArea1'	Length of building (m)	13	5127.432
'FloorArea2'	Width of building (m)	12	5128.906
'FloorHeight'	Height of building (m)	7	29987.17
'WindowArea'	Window to wall ratio	1	65185.2
'ZoneUnitLW'	Length and Width of zone (m)	3	47566.7
'R_Wall'	R Value of Wall (m2 K/W)	15	2517.623
'R_Roof'	R Value of Roof (m2 K/W)	16	2065.674
'U_Window'	U value of window (W/m2 K)	8	27407.89
'VentNumPeople'	Number of People	5	32283.65
'VentRate'	Ventilation rate of one person	6	31723.18
'Eff'	Efficiency of boiler	4	35456.67
'ThickWall'	Thickness of wall (m)	9	22925.24
'ThickRoof'	Thickness of roof (m)	21	0.422604
'ThickFloor'	Thickness of floor (m)	18	3.107767
'Tzone'	Zone temperature (°C)	2	55840.39
'density_wall'	Wall density (kg/m3)	14	4104.264
'Cp_wall'	Capacity of wall (J/kg-C)	11	6114.315
'density_roof'	Roof density (kg/m3)	26	0.080669
'Cp_roof'	Capacity of roof (J/kg-C)	25	0.119276
'density_floor'	Floor density (kg/m3)	22	0.411091
'Cp_floor'	Capacity of floor (J/kg-C)	20	0.605462
'hcf'	Floor convection coefficient	24	0.330991
'hcc'	Ceiling convection coefficient	23	0.386241
'hcw'	Wall convection coefficient	19	1.108791
'hco'	Outdoor convection coefficient	10	7655.817
'alpha'	Solar absorptance coefficient of wall surface for solar radiation	17	631.9552
'emit'	Effective emittance	27	0

4.3.3. Description of the case study

To validate the proposed methodology, a RC model for a building was created and tested in a closed-loop situation. Table 4-5 represents the input parameters for the reference data and ranges.

Table 4-5. Default input parameters for RC model

Variable Name	Description	Default Value	Minimum value	Maximum value
'FloorArea1'	Length of building (m)	52.5	5	100
'FloorArea2'	Width of building (m)	52.5	5	100
'FloorHeight'	Height of building (m)	15	10	20
'WindowArea'	Window to wall ratio	0.5	0.1	0.9
'ZoneUnitLW'	Length and Width of zone (m)	16	12	20
'R_Wall'	R Value of Wall (m2 K/W)	27.5	5	50
'R_Roof'	R Value of Roof (m2 K/W)	27.5	5	50
'U_Window'	U value of window (W/m2 K)	0.735	0.5	0.97
'VentNumPeople'	Number of People	8.5	2	15
'VentRate'	Ventilation rate of one person	17.5	5	30
'Eff'	Efficiency of boiler	0.74	0.5	0.98
'ThickWall'	Thickness of wall (m)	0.6	0.2	1
'ThickRoof'	Thickness of roof (m)	0.6	0.2	1
'ThickFloor'	Thickness of floor (m)	0.6	0.2	1
'Tzone'	Zone temperature (°C)	22.5	15	30
'density_wall'	Wall density (kg/m3)	2250	2000	2500
'Cp_wall'	Capacity of wall (J/kg-C)	600	500	700
'density_roof'	Roof density (kg/m3)	2250	2000	2500
'Cp_roof'	Capacity of roof (J/kg-C)	600	500	700
'density_floor'	Floor density (kg/m3)	2250	2000	2500
'Cp_floor'	Capacity of floor (J/kg-C)	600	500	700
'hcf'	Floor convection coefficient	5	4	6
'hcc'	Ceiling convection coefficient	5	4	6
'hcw'	Wall convection coefficient	5	4	6
'hco'	Outdoor convection coefficient	5	4	6
'alpha'	Solar absorptance coefficient of wall surface for solar radiation	0.4	0.5	0.8
'emit'	Effective emittance	0.3	0.2	0.4

These case studies utilized a personal computer with an AMD FX-6200 (Benchmark score: 6,145 as of 3 Jun 2014) (CPU). CPU Benchmark results (“Baselines”) were gathered from users’ submissions to the PassMark web site as well as from internal testing. The PerformanceTest conducts eight different tests and then averages the results together to determine the CPU Mark rating for a system. The table below provides examples of commonly used CPUs.

Table 4-6. Passmark score of commonly used CPUs (CPU)

CPU Name	Passmark CPU Mark (Higher is better)	Rank (Lower is better)	CPU Value (Higher is better)
AMD FX-6200 Six-Core	6142	190	42.36
Intel Core2 Duo E4400@ 2.00Ghz	1122	1192	10.39
Intel Core i7-4770K@ 3.50Ghz	10328	32	32.28
Intel Pentium 4 3.00Ghz	362	1671	NA

The weather data that was used in these sensitivity cases were from Golden, CO.

4.3.4. Sensitivity analysis

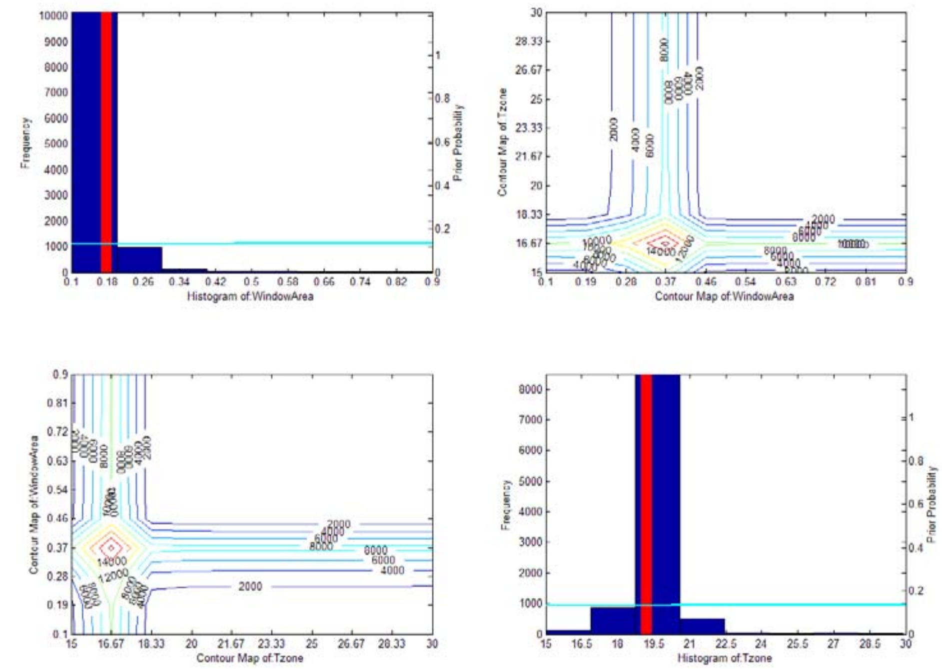
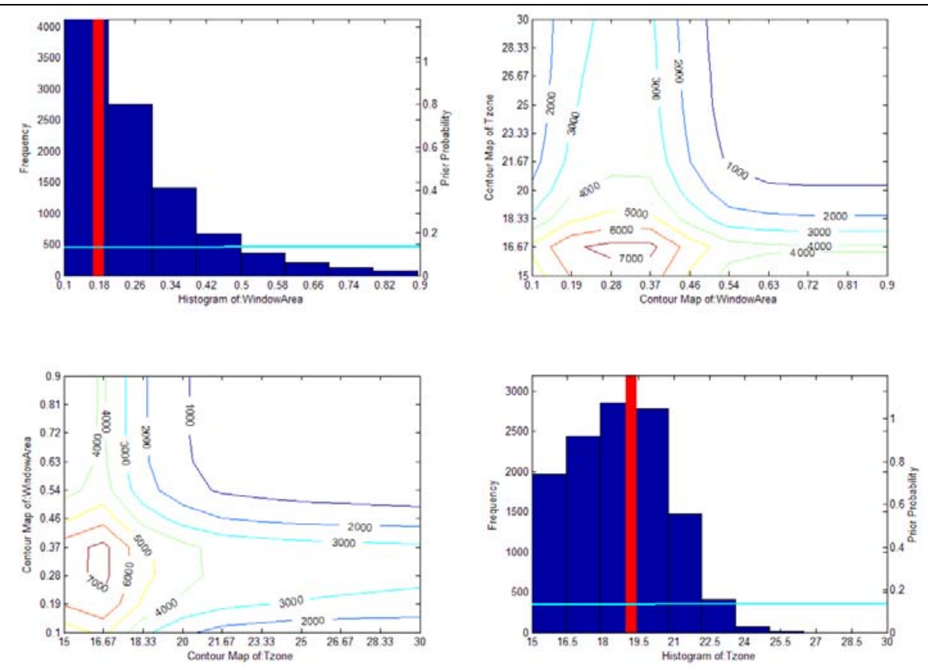
In order to perform Bayesian methodology for unknown parameter identification, it is crucial to identify those parameters within the Gaussian process or Bayesian approach. These parameters can be ambiguous due to the fact that there are no absolute values for these parameters. These parameters include sampling number, observation noise, number of unknown parameters that can be identified with this approach, etc. In this section, validation of these parameters will be performed in order to search for appropriate value for these parameters. In a Bayesian framework, especially if this framework will be utilized in the reality by normal users, two important aspects must be kept in mind: time consumption and accuracy. If the process requires 40 hours or more to run the framework, accessibility is severely limited. Similarly, it is a significant impairment if the framework generates results with very low accuracy. Therefore, it is necessary to conduct sensitivity analysis.

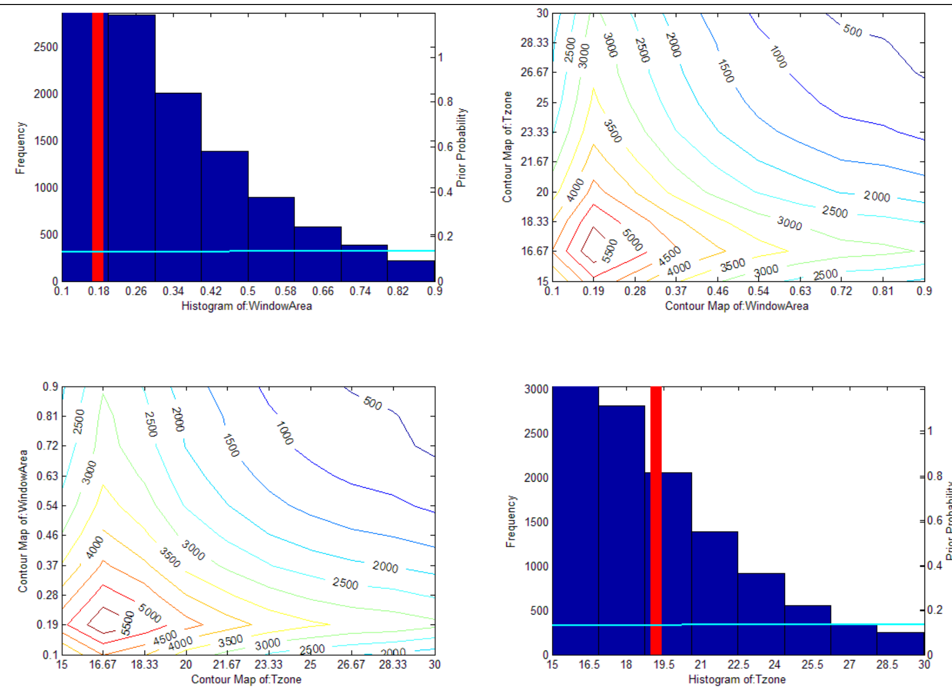
4.3.4.1. Observation noise

The observation noise (σ_y) is used in a likelihood calculation when performing MCMC sampling. This value determines how likely the simulated or emulated data is close to the reference data. Therefore, it is important to have an appropriate value since this will have an impact on the acceptance rate during MCMC sampling. Acceptance rate is the fraction of proposed samples that are accepted in a total period of sampling numbers. There is no specific criterion that specifies acceptance rate, but Gelman (Gelman, et al., 2013) suggested that the acceptance rate should be 45% for uni-dimensional problems and 25% for six-dimensional problems.

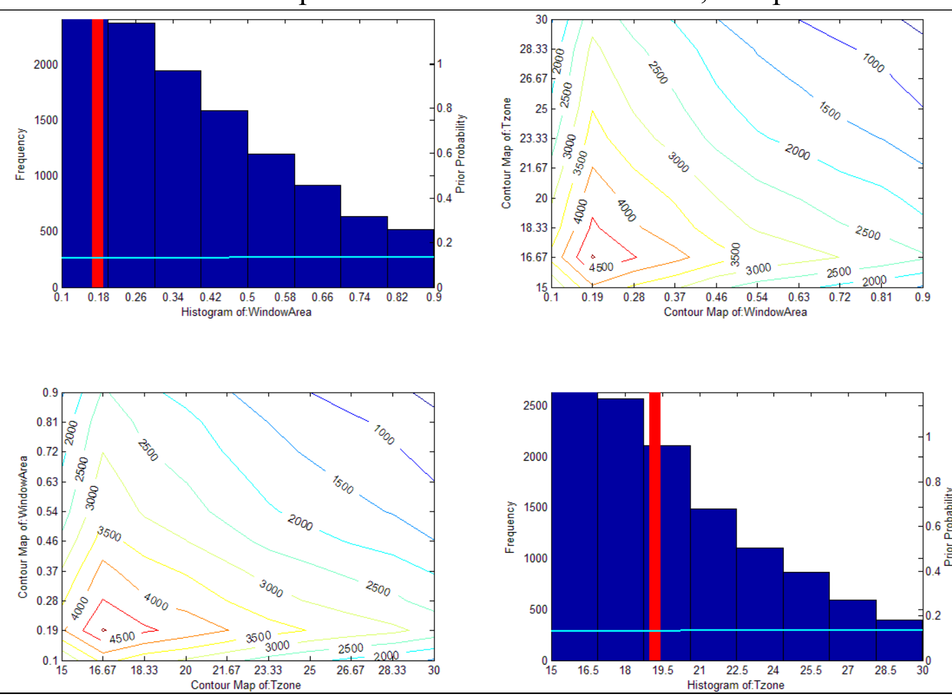
In this case study, we have performed a sensitivity analysis on several values of observation noise. Table 4-7 presents those results. In the table, the red bar line in the figure represents the target reference value and the graph itself is the posterior distribution after MCMC sampling. Lastly, the contour graphs are the posterior result of each parameter when paired with other parameters. The sampling number that was used in this case was 12,000. As shown in the table, all of observation errors seem to identify the correct reference values that this approach was seeking. However, the acceptance rate of each case study tended to increase while observation error increased exponentially. For the case of 2 unknown parameter identification case, if the observation error is more than $\sigma_y = e^{-6.5}$, the acceptance rate increases and thus occasionally fails to identify the correct value. An observation error of less than $\sigma_y = e^{-6.5}$ tends to correctly identify the parameter, but the result show an extremely high posterior distribution. This can be interpreted as lower observation error that tends to have a lower acceptance rate, but fails to search the wide range of parameters regardless of the prior belief of the parameters. As Gelman suggested, a reasonable acceptance rate should be between 25%~ 45%. As a result, this study proposes that an observation noise of $\sigma_y = e^{-6.5}$ should be chosen for a two unknown parameter case and $\sigma_y = e^{-4}$ for a five unknown parameter case to be used.

Table 4-7. Sensitivity result of observation error in RC model

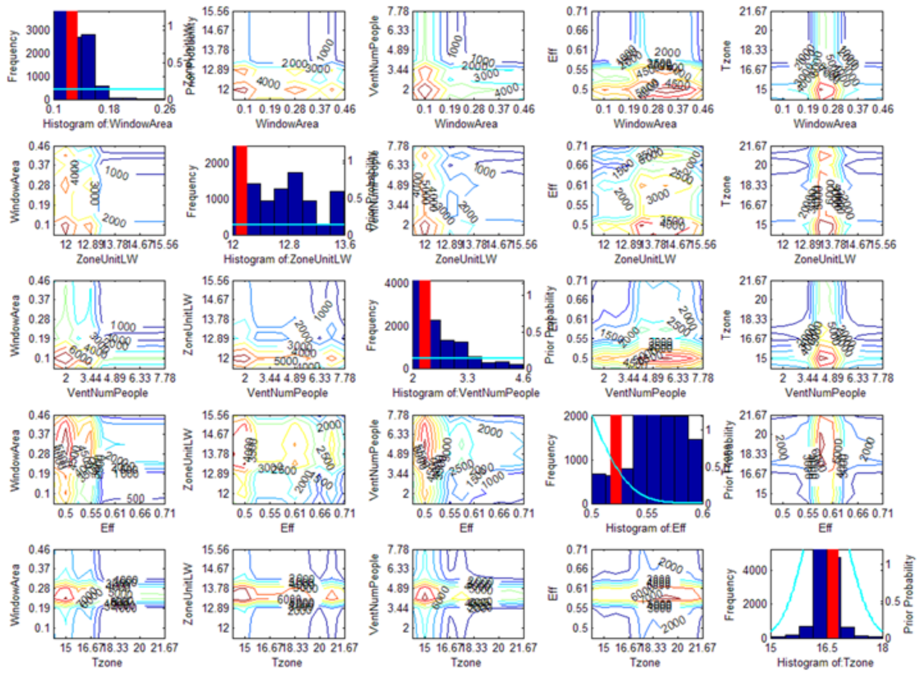

$\sigma_y = e^{-10.5}$ with 2 unknown parameters - CPU Time: 39 sec, acceptance rate: 5%

$\sigma_y = e^{-6.5}$ with 2 unknown parameters - CPU Time: 39 sec, acceptance rate: 21%



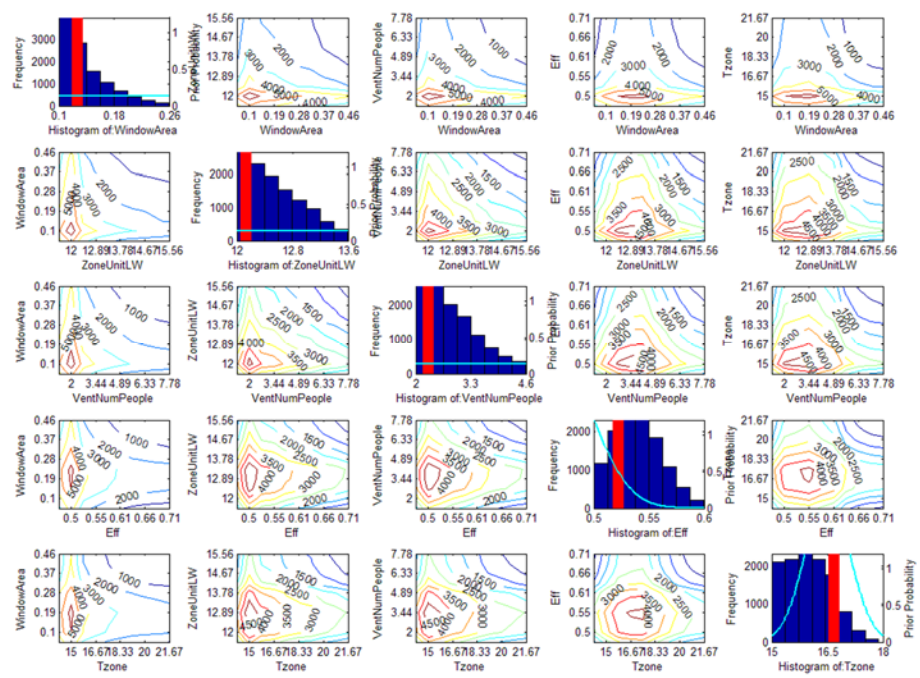
$\sigma_y = e^{-4}$ with 2 unknown parameters - CPU Time: 39 sec, acceptance rate: 51%



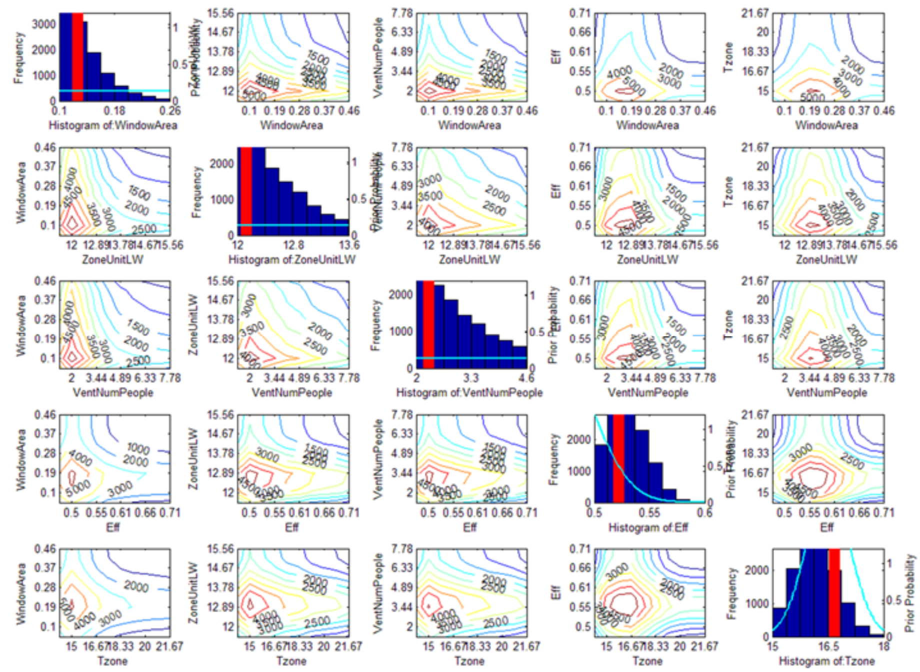
$\sigma_y = e^{-2.5}$ with 2 unknown parameters - CPU Time: 39 sec, acceptance rate: 66%



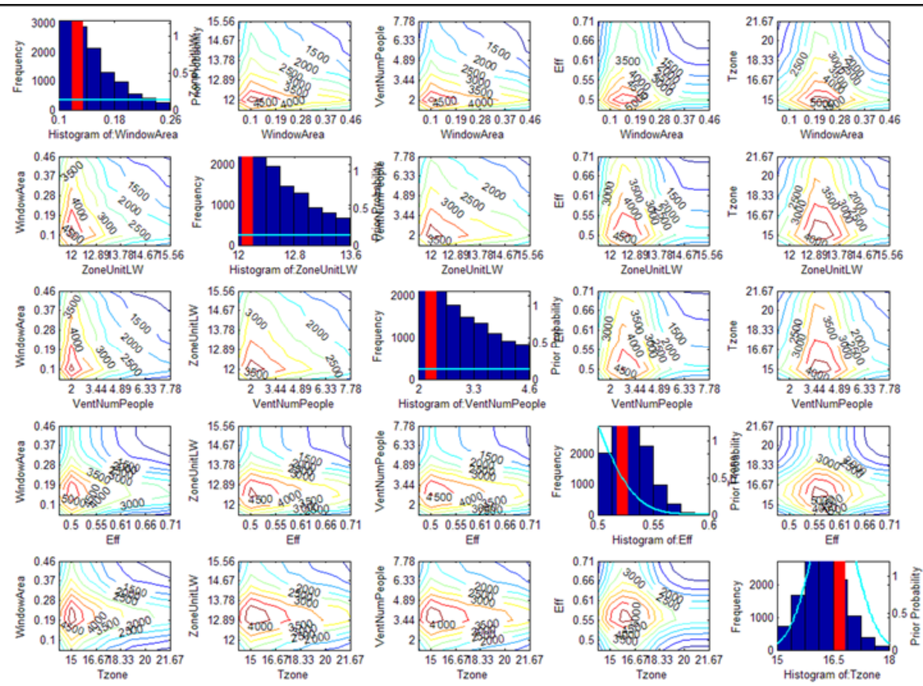
$\sigma_y = e^{-8}$ with 5 unknown parameters - CPU Time: 107 sec, acceptance rate: 8%



$\sigma_y = e^{-4}$ with 5 unknown parameters - CPU Time: 108 sec, acceptance rate: 31%



$\sigma_y = e^{-2}$ with 5 unknown parameters - CPU Time: 108 sec, acceptance rate: 54%



$\sigma_y = e^{-1}$ with 5 unknown parameters - CPU Time: 108 sec, acceptance rate: 62%

4.3.4.2. *Size of Gaussian process model and number of unknown parameters*

The size of an emulator (Gaussian process regression model) can impact accuracy. The size of an emulator means the size of training points for the Gaussian process. For instance, 1,000 training points generates a 1,000 by 1,000 covariance matrix. In this sensitivity analysis, this study generated several cases (e.g., 100 training points, 500 training points, 1,000 training points) to verify the accuracy of the model. In comparison, the annual percentage error (APE) score was used to calculate how close the simulation outcomes are to the reference data. In this case study, 1,000 reference data (1,000 testing points) were generated by Latin hypercube sampling. The below equation explains how an APE score is calculated. Utility data was replaced by testing points.

$$APE = \sqrt{\sum_{i=1}^{12}\left(\left[\frac{Elec_{i,util} - Elec_{i,sim}}{Elec_{yr,util}}\right]^2 + \left[\frac{Gas_{i,util} - Gas_{i,sim}}{Gas_{yr,util}}\right]^2\right)} \tag{4.7}$$

Where:

$Elec_{i,util}$ = Monthly electricity energy usage of utility data

$Elec_{i,sim}$ = Monthly electricity energy usage of simulation data

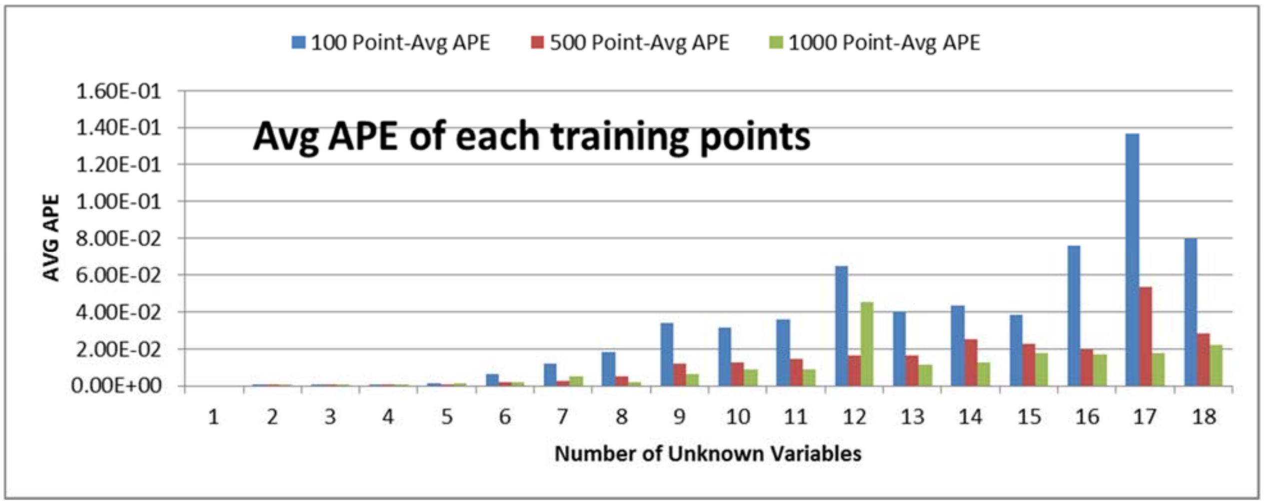
$Elec_{yr,util}$ = Yearly electricity energy usage of utility data

$Gas_{i,util}$ = Monthly Gas energy usage of utility data

$Gas_{i,util}$ = Monthly Gas energy usage of simulation data

$Gas_{i,util}$ = Yearly Gas energy usage of utility data

Figure 4-10 represents the maximum APE score and CPU time for 100, 500, 1,000 training points when compared with 1,000 testing points. As shown in the figure, 100 training points tend to have a larger APE score and 1,000 training points had the lowest APE score. A maximum APE score is shown in the figure. Although an average APE score can reside in the reasonable tolerance range, it is important to check the maximum APE score. Checking this maximum APE minimizes the difference between a simulator and emulator to within a certain range. We want an emulator that generates reasonable value within a certain range of parameters. As such, it is necessary to consider a maximum APE score. The results showed that reasonable outputs are generated when compared to the simulator for up to 16 unknown parameters.



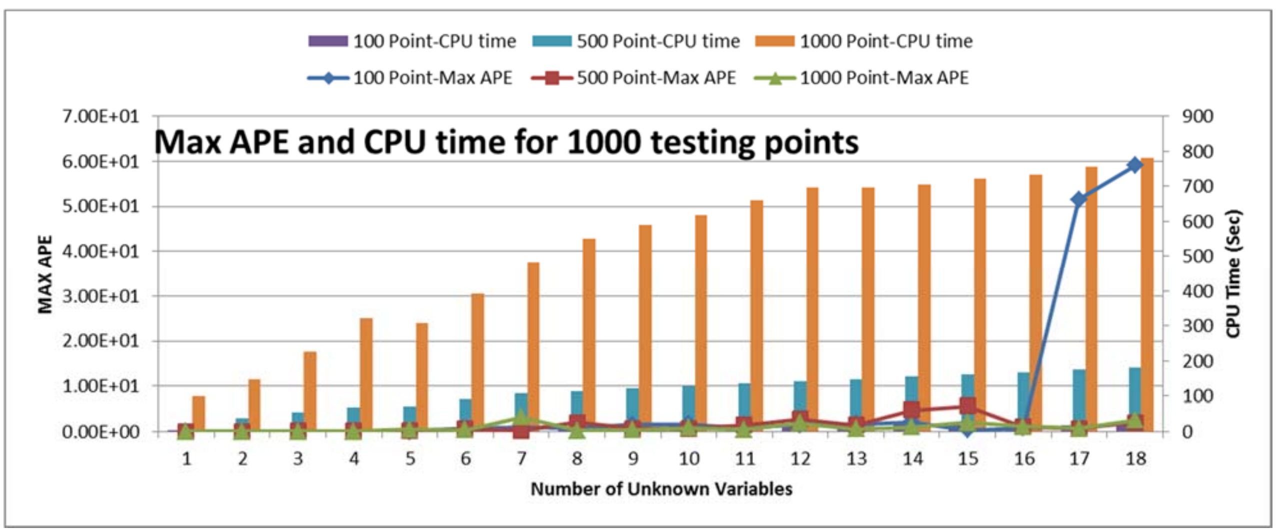


Figure 4-10. Maximum, average APE and CPU time for 1000 testing points

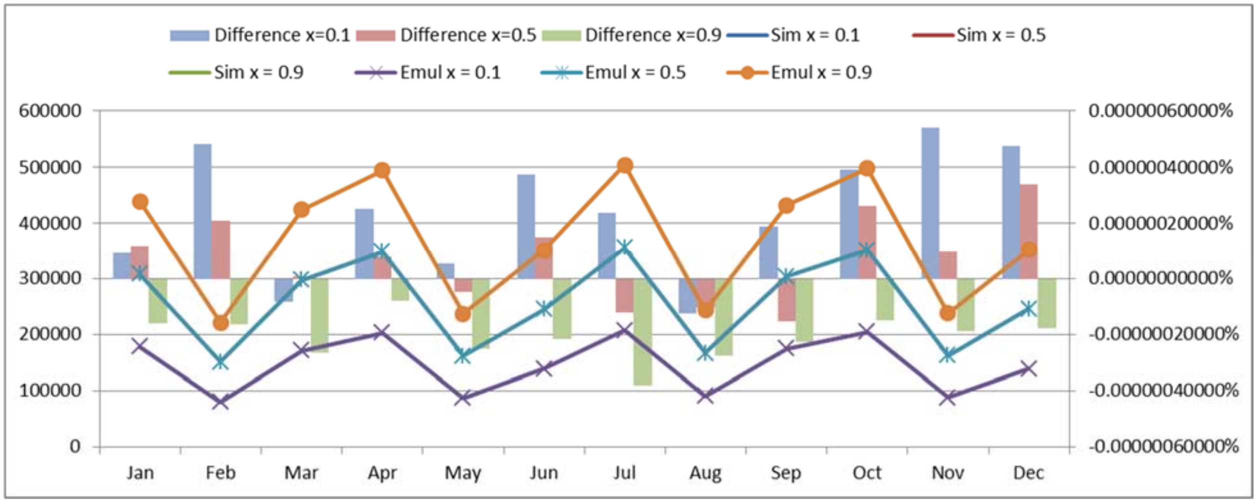


Figure 4-11. Emulator result compared with simulator

Figure 4-11 represents the monthly energy consumption result of an emulator and simulator in a case with one unknown parameter identification. The training point that was used for this case was 1,000. As shown in the figure, it is difficult to observe any difference between the emulator and simulator. The right Y-axis and bar graph represent the difference between these two engines.

4.3.4.3. Continuous variable and discrete variable

As mentioned previously, there are various parameters involved in building energy analysis. Some of the parameters have value characteristics that can be defined as continuous, such as the R-value of materials. This means that the parameter can range continuously from one value to another. This is not the case, however, for some parameters. For instance, window glass type cannot be chosen in a continuous range. For the proposed Bayesian approach, discrete parameters are required to be evaluated. Additionally, to reduce the computing burden and to reduce the searchable range for parameters, discrete variable need to be considered.

This study performed parameter identification of discrete variables. Figure 4-12 presents two parameter identification results. The right-hand side shows the continuous parameter results and the left plot represents discrete parameter results. Discrete parameter identification gave a narrower result in the contour plot than the continuous results. This is because there is less search needed than that for continuous parameter identification. The results for acceptance in both analyses were similar at 23%. This means that even though there is no benefit in terms of computing time, discrete parameter identification can result in more accuracy due to a reduced search field.

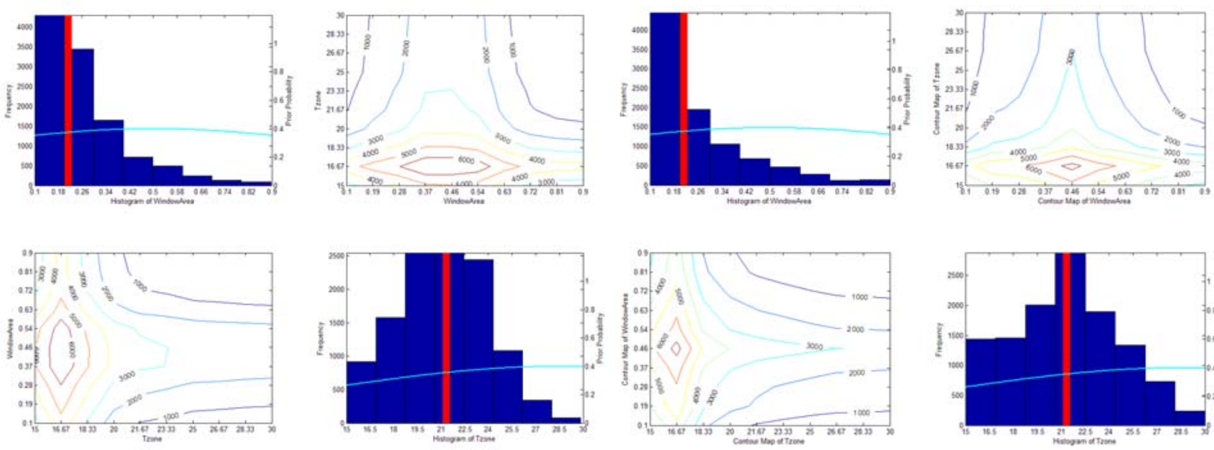


Figure 4-12. Comparison of continuous and discrete parameter identification in RC model

4.4. Validation of DOE2.2 model

4.4.1. Introduction

DOE2.2 (DOE2) is free software and is widely used as a building energy analysis program to predict energy use and cost for various types of buildings. Along with energy use and cost, it analyzes almost all energy-related fields that apply to a building. It uses the geometry of a building, construction, operating schedules and system related information along with a location weather file. DOE2.2 computes an hourly simulation of the building and provides estimates in an energy analysis. eQUEST is a graphic user interface for DOE2.2 and it provides wizards and graphic displays for detailed comparative analysis of building designs and technologies.

In this study, we utilized the DOE2.2 model to validate our Bayesian framework in order to verify whether the framework can function as an hourly simulation tool. Unlike EnergyPlus, DOE2.2 is known to be a light (in terms of computer burden) tool that is capable of providing accurate results. Rallapalli (RallapalliHema, 2010) performed energy analysis on a medium office building with both EnergyPlus and eQUEST and found out that there was a -1% difference in annual electricity consumption. Their study also stated that EnergyPlus can be more accurate than DOE2.2, however, EnergyPlus took longer to compute. Therefore in this study, DOE2.2 will be utilized to accommodate a Bayesian framework.

4.4.2. Simple DOE2.2 model

4.4.2.1. Description of simple DOE2 model

In order to verify validity for the parameter identification, several case studies were done. Utility data was created by DOE2.2 engine with variable settings as shown in the below table. The given case uses gas domestic hot water heater and furnace with dx cooling. Remaining main input values are shown in Table 4-8. The overall building shape and appearance is shown in Figure 4-13.

Table 4-8. Other Input setting for DOE2.2 modeling

Parameter name	Value
Width of the building	64 ft
Length of the building	80 ft
Floor height of the building	12 ft
WWR	25%
R value for the wall	10 ft^2 F hr / Btu
R value for the roof	25 ft^2 F hr / Btu
U value for the window	0.7 Btu / ft^2 F hr
Number of People per zone	8
Vent Rate	20 cfm per person
Boiler Efficiency	0.85
Thickness of wall	0.5 ft
Thickness of roof	0.5 ft
Thickness of floor	0.5 ft
Set temperature for the zone	20 C
Weather file location	CO_Greeley_Weld_(AWOS)

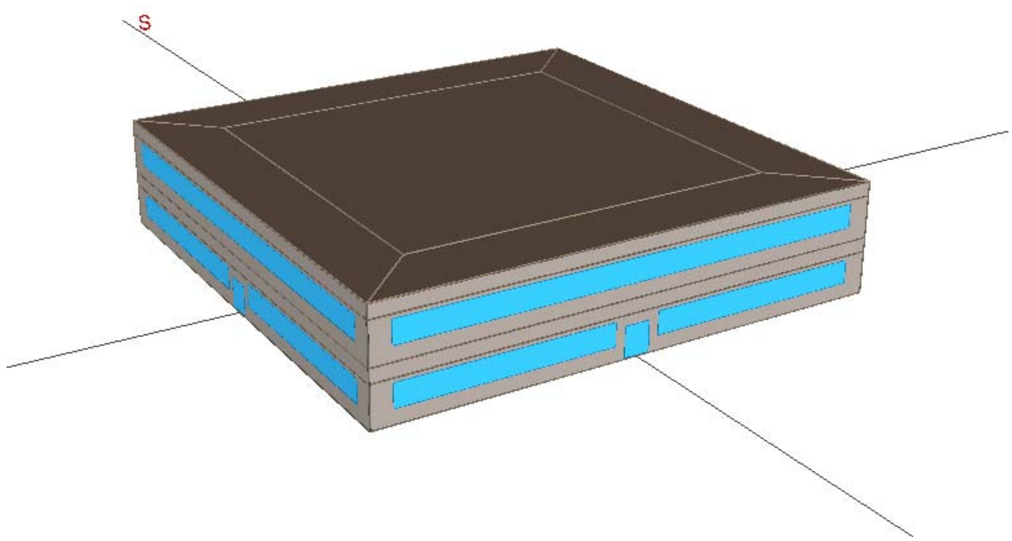


Figure 4-13. Example building's appearances of DOE2.2 model

4.4.2.2. Sensitivity analysis of simple DOE2.2 model parameters

To validate the proposed Bayesian framework, a DOE2.2 model of a building was created and tested in a closed-loop situation for RC model verification. As mentioned previously, a 'closed loop situation' is used instead of real-time data in order to achieve simplicity and accuracy for model validation. In order to do this, we obtained DOE2.2's reference data by using a preset value of parameters before going into Bayesian analysis. With the obtained reference data, we can set all or a partial set of the parameters as unknown and try to identify these unknown parameters.

Before we go into further analysis, sensitivity analysis of these unknown parameters is needed in order to accommodate sensitivities for each parameter. Each parameter has a different influence on energy consumption. Further, some parameters have a negligible influence on the energy consumption and thus can be ignored in the tuning process for the energy modeling process. A normal procedure for intuitive calibration starts with parameters that have a larger influence on energy consumption than parameters with a lower influence.

In DOE2.2 energy modeling there are thousands of input parameters (depending on the complexity of a building) that need to be selected by the user. Among that large number of input parameters, this study chose 10 variables to be varied within a Bayesian framework as shown in Table 4-9. This sensitivity result was calculated based on the mean value of each parameter, and electricity and gas energy consumption, with a result of 10% variation for each. The results showed that variation in infiltration had the largest energy consumption difference. However, compared to the results for other variables, the 10% variation result for infiltration was an extreme result. As such, the validation study will be processed excluding this parameter. In a 10 parameter identification case, however, this parameter will be included.

Table 4-9. Mean, Min, Max value and Sensitivity analysis result of input parameters for DOE2.2 model

Variable Name	Description	Mean Value	Min Value	Max Value	Rank	Impact on Energy Consumption (mBtu) per 10% variance
'RoofResistance'	Roof Resistance (h-ft2-F/Btu)	18.5	5	32	8	0.794
'WallResistance'	Wall Resistance (h-ft2-F/Btu)	18.5	5	32	7	1.066
'WindShadeCo'	Window Shade Coefficient	1.05	0.1	2	2	6.320
'Infil'	Infiltration (CFM/ft2)	1.255	0.5	2.01	1	85.911
'OccuDens'	Occupancy Density (ft2/person)	140	20	260	10	0.682
'LightDens'	Light Density (W/ft2)	1.9	0.5	3.3	5	4.181
'EquipDens'	Equipment Density (W/ft2)	1.9	0.5	3.3	3	5.123
'DHWHIR'	Heat Input Ratio of Domestic Hot Water	1.8	0.5	3.1	4	4.503
'CoolingEIR'	Energy Input Ratio of Cooling	0.45	0.1	0.8	9	0.740
'HeatingHIR'	Heat Input Ratio of Heating	1.85	0.1	1.6	6	2.101

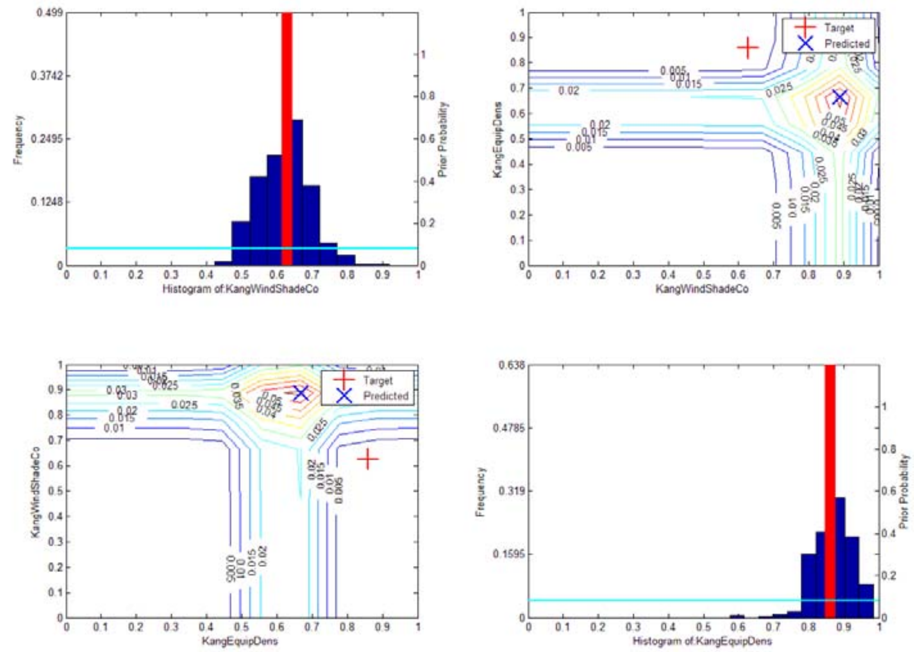
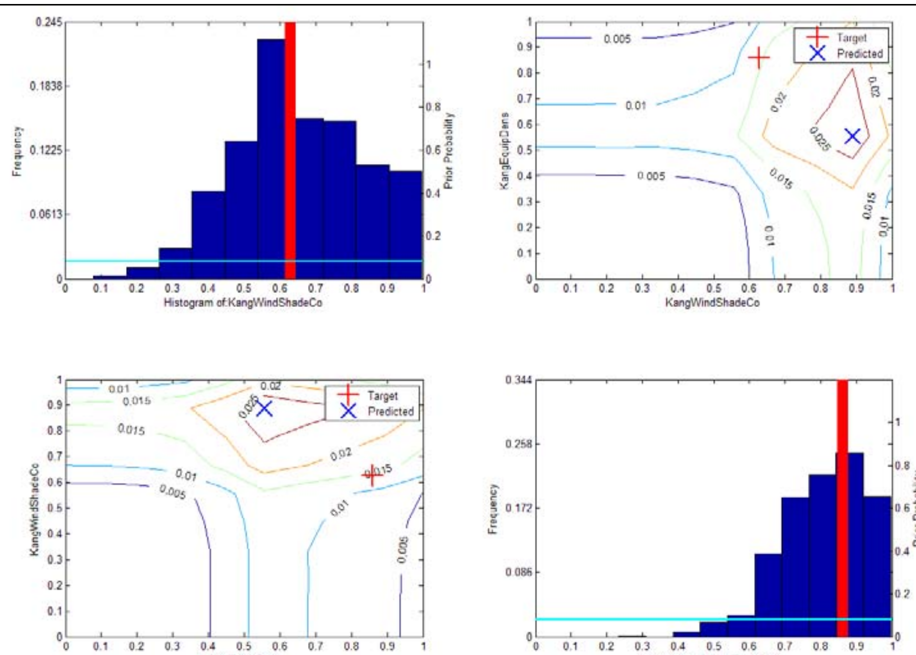
4.4.2.3. Sensitivity analysis

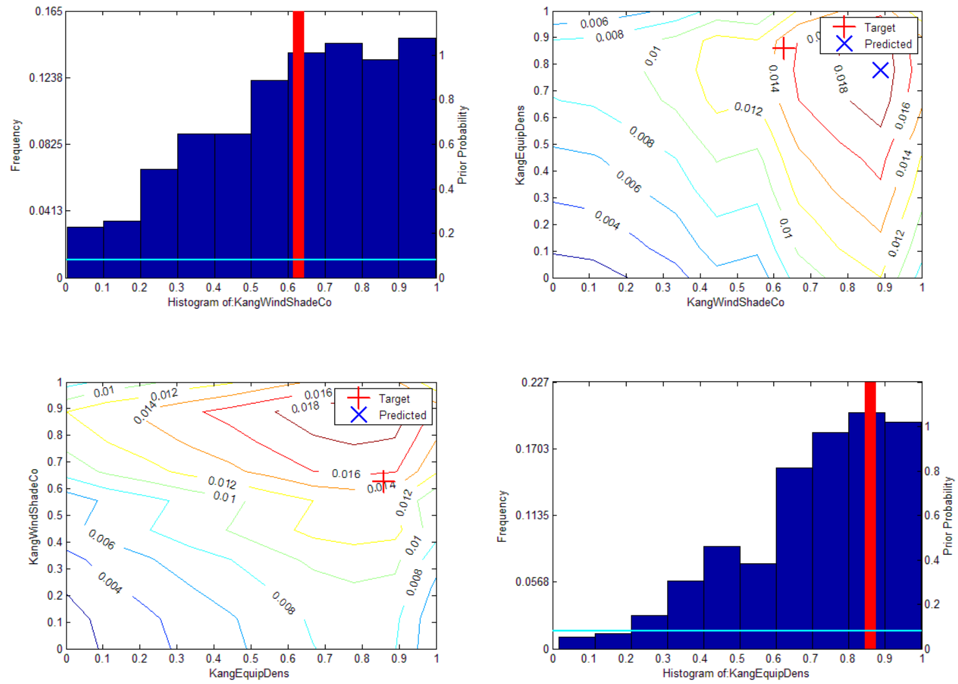
4.4.2.3.1. Observation noise

In this case study, the sensitivity analysis was done for several values of observation noise. Table 4-10 presents the results. As with the sensitivity results in the RC model, a red bar line in the diagonal plot within each figure represents the target reference value that we are looking for and the graph itself is the posterior distribution after MCMC sampling. Lastly, the contour graphs are the posterior result of each parameter when paired with other parameters. A plus marker '+' represents the predicted target value from the Bayesian framework and the X marker 'X' represents the original target value. The sampling number used was 1,000. As shown in the table, all the observation errors seem to identify a correct reference value that this approach was seeking, similar to the RC model cases. The acceptance rate of each case study tended to increase while observation error increased.

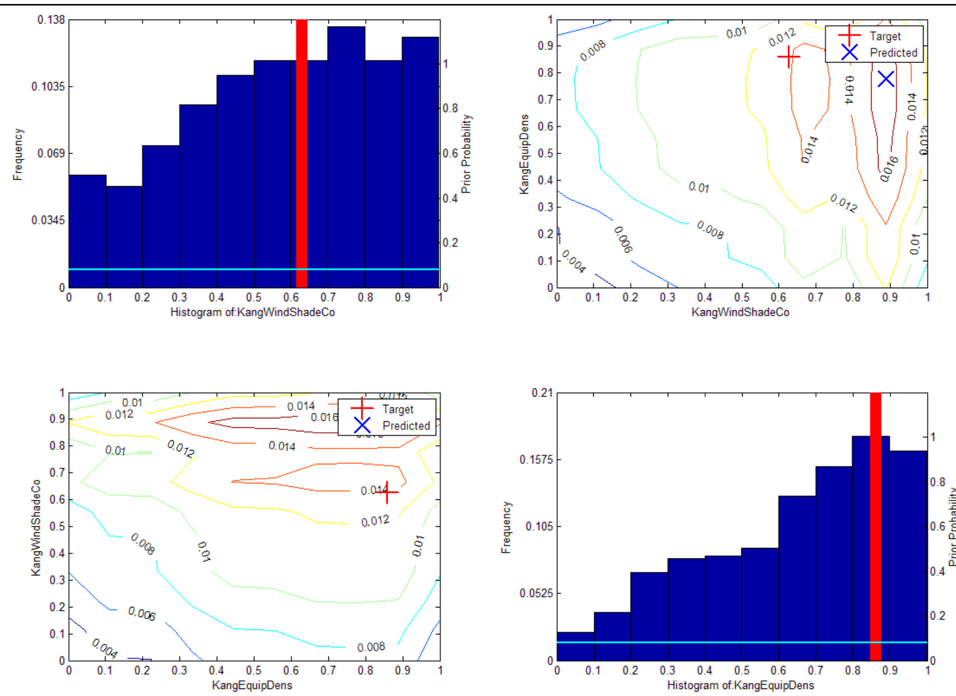
For five unknown parameters, this study applied a prior distribution around the target value since the Bayesian framework kept failing to find a target value when applied with a uniform prior distribution. As shown in the table, if the observation error is more or less than $\sigma_y = e^{-7.5}$ (two unknown parameter case) and $\sigma_y = e^{-8}$ (five unknown parameter case), despite of pointing to the right target value, the acceptance rate tends to increase or decrease following the observation's increase or decrease. Like RC model, since the acceptance rate should be to fit in 25% ~ 45%, the observation error of $\sigma_y = e^{-7.5}$ (two unknown parameter case) and $\sigma_y = e^{-8}$ (five unknown parameter case) should be chosen.

Table 4-10. Sensitivity result of the observation noise in simple DOE2.2 model

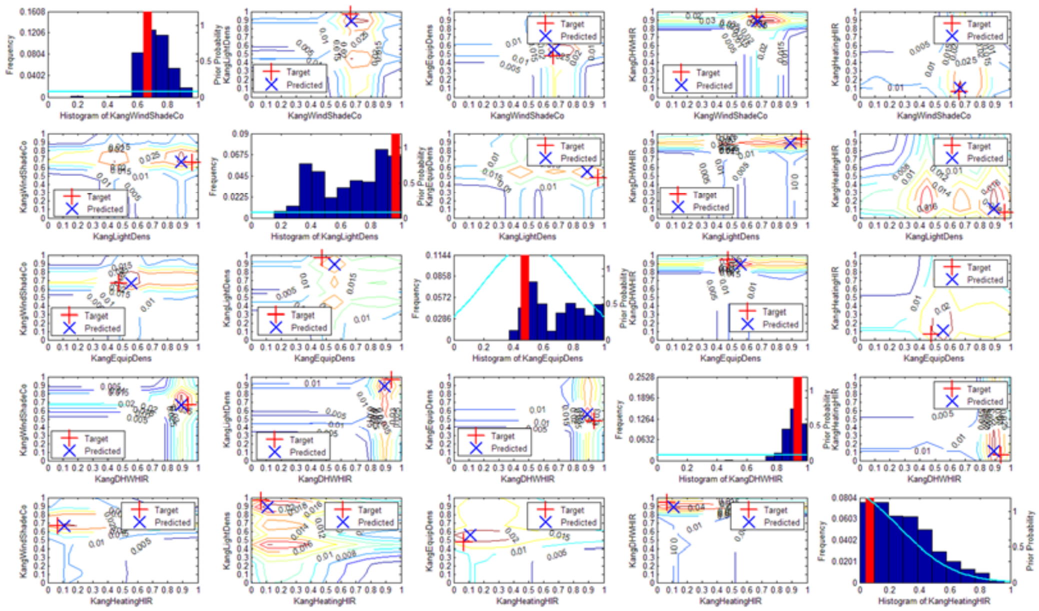

$\sigma_y = e^{-9.5}$ with 2 unknown parameters - CPU Time: 76.4 sec, acceptance rate: 15%

$\sigma_y = e^{-7.5}$ with 2 unknown parameters - CPU Time: 75 sec, acceptance rate: 42%



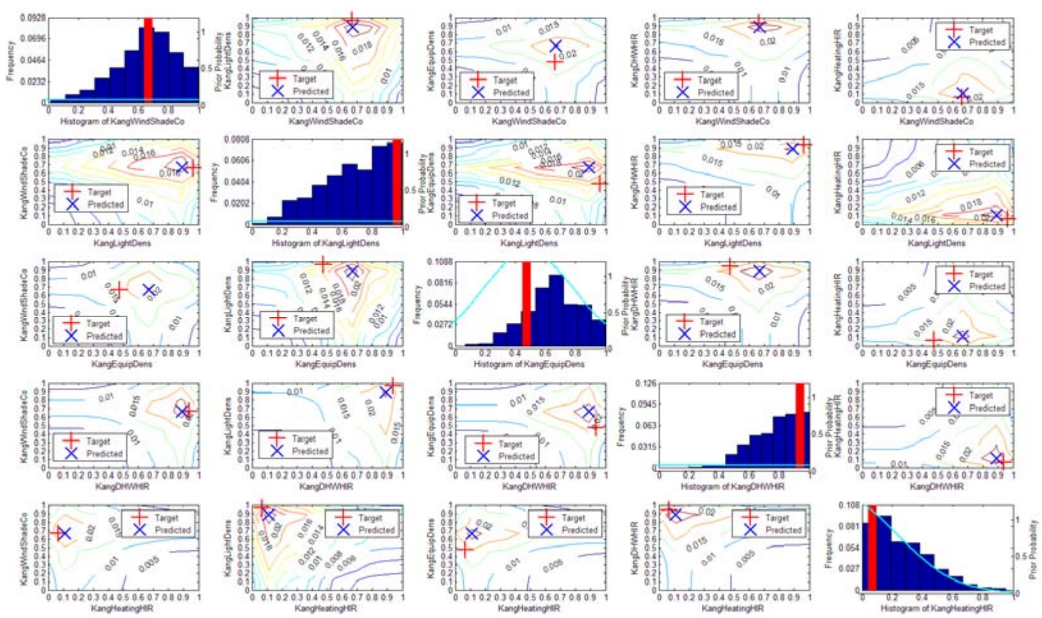
$\sigma_y = e^{-6.5}$ with 2 unknown parameters - CPU Time: 73.8 sec, acceptance rate: 68%



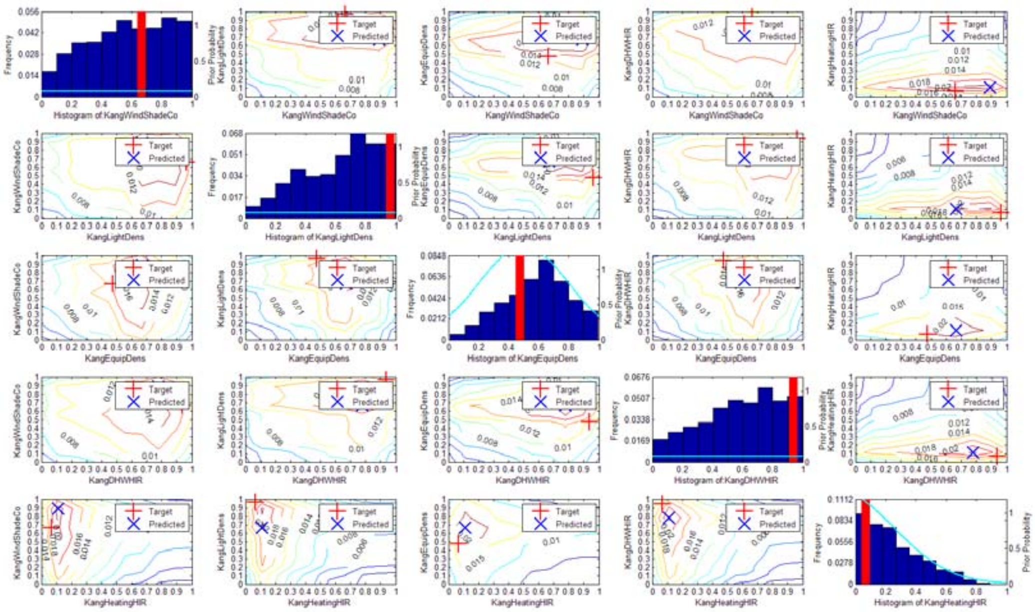
$\sigma_y = e^{-6}$ with 2 unknown parameters - CPU Time: 72.4 sec, acceptance rate: 75%



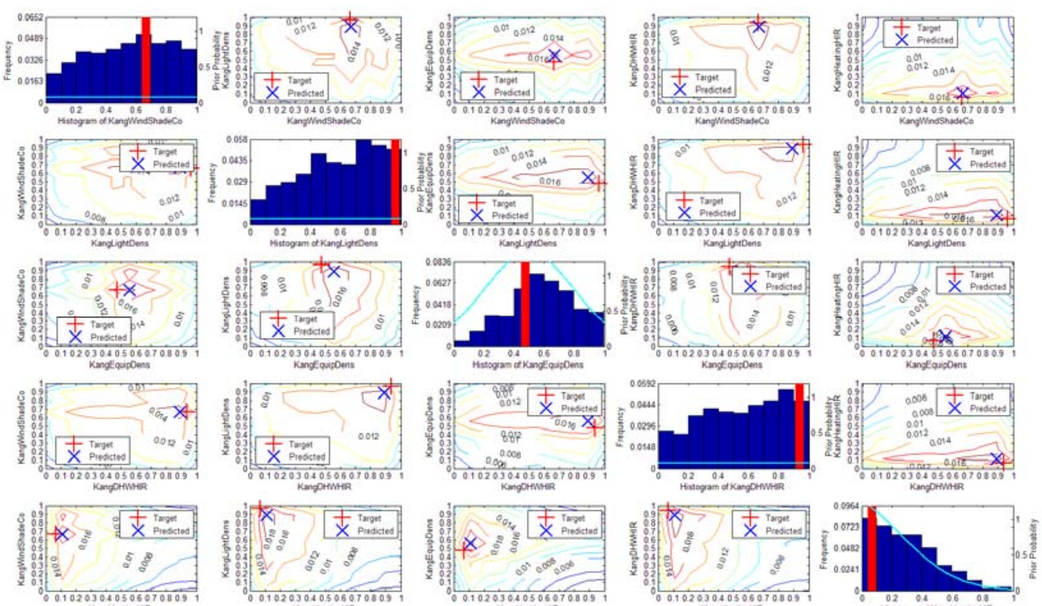
$\sigma_y = e^{-10}$ with 5 unknown parameters - CPU Time: 106.9 sec, acceptance rate: 23%



$\sigma_y = e^{-8}$ with 5 unknown parameters - CPU Time: 106.9 sec, acceptance rate: 47%



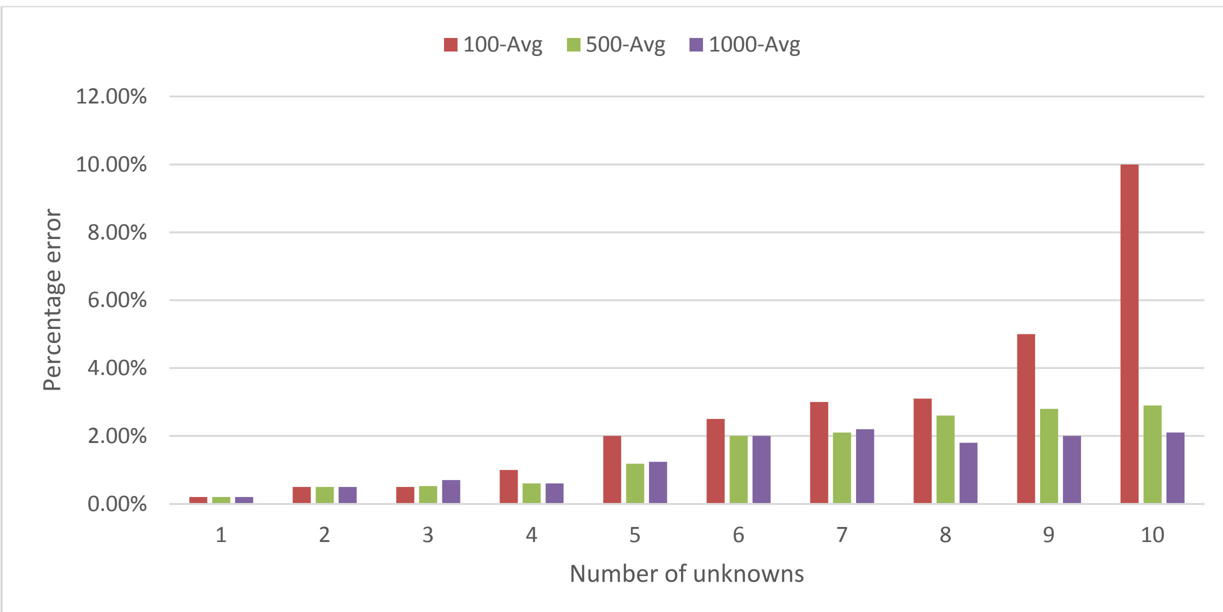
$\sigma_y = e^{-7}$ with 5 unknown parameters - CPU Time: 114.8 sec, acceptance rate: 67%



$\sigma_y = e^{-6.5}$ with 5 unknown parameters - CPU Time: 115.4 sec, acceptance rate: 76%

4.4.2.3.2. Size of Gaussian process

Size of Gaussian process is important to determine the emulator's accuracy and it has big influence on CPU time of the whole process. As the size of a Gaussian process increases, the CPU time might increase exponentially due to the fact that Gaussian process requires inverse of huge matrices. Therefore, the size of the Gaussian process is the main subject in this case study. In Figure 4-14 represents the average accuracy and errors that are over 10% are shown in the table along with its CPU time. Three cases were selected as a training size of Gaussian process. As shown in the figure, error tends to decrease as size of Gaussian process increases.



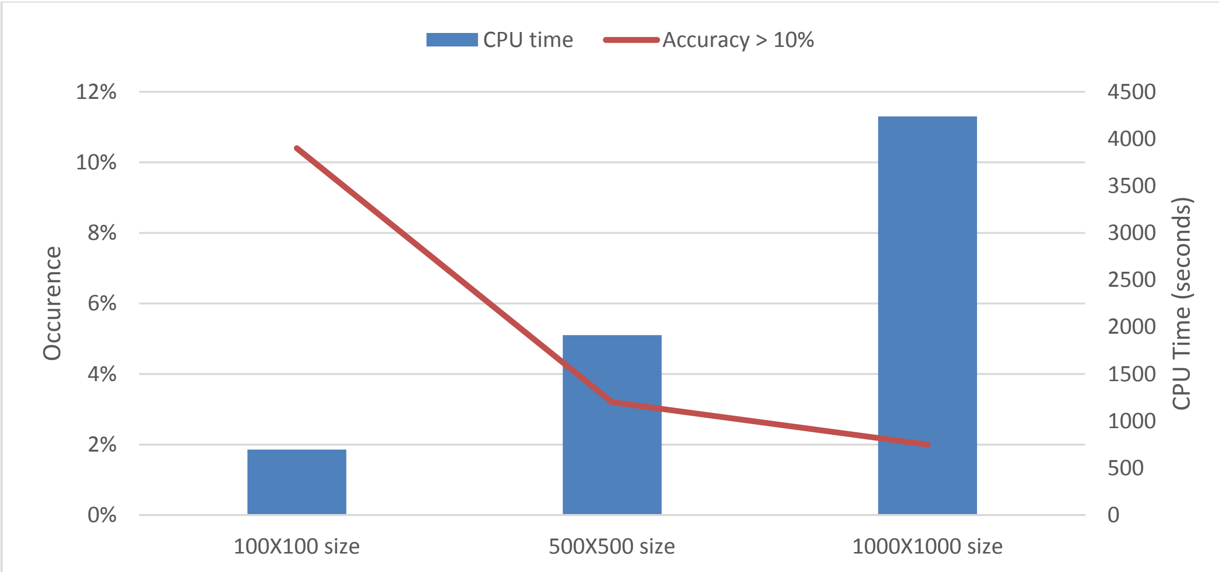
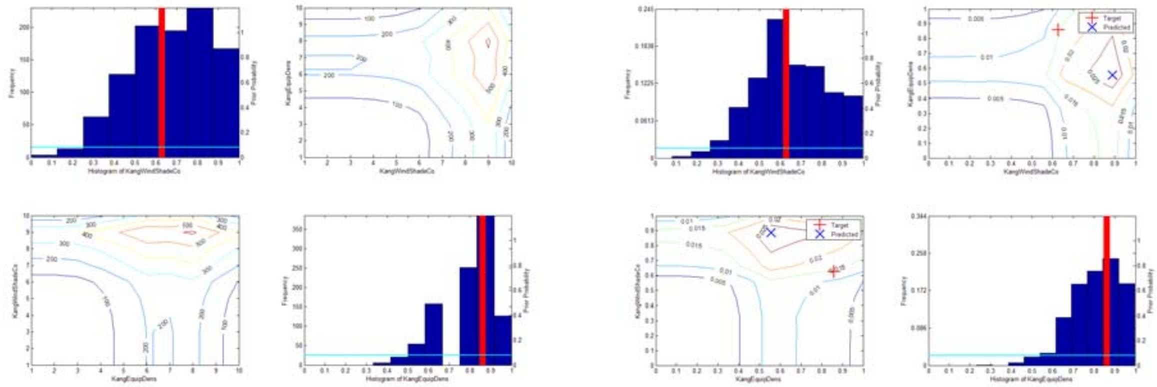


Figure 4-14. Size of Gaussian process results in Simple DOE2.2 model – average percentage error (top) and occurrences of error over 10% (bottom)

4.4.2.3.3. Continuous variables and discrete variables

In this study, 2 unknown parameter identification case for DOE2.2 with sampling number of 1,000 is performed. As shown in Figure 4-15, both cases seem to find the appropriate target value despite of the variable being “Discrete” or “Continuous.” As shown in the figure, there is little difference when finding the target value.



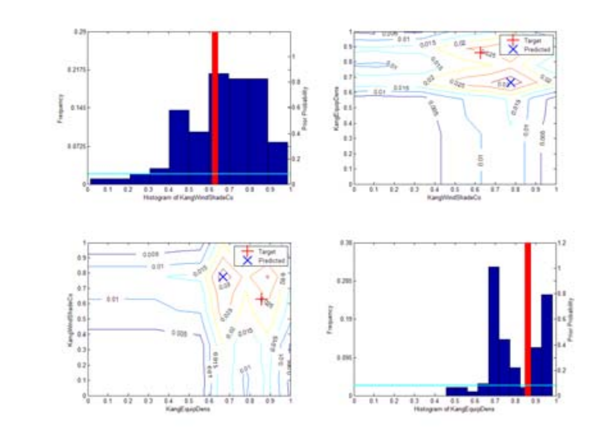
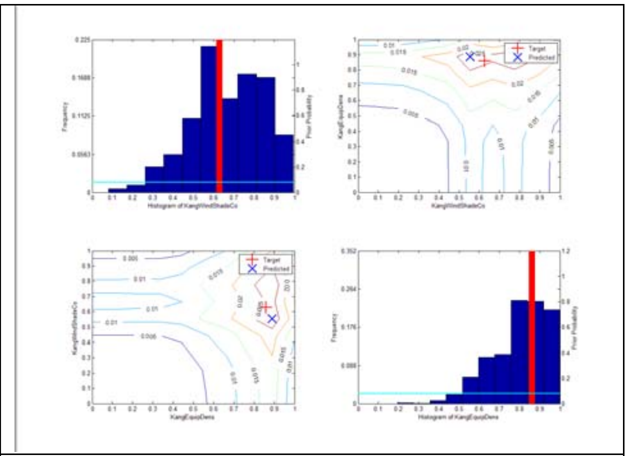
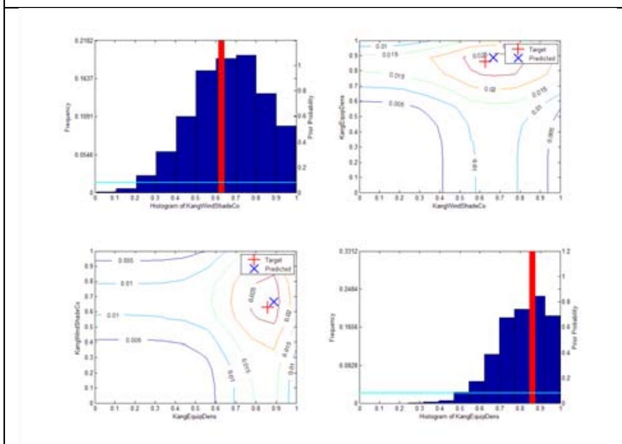
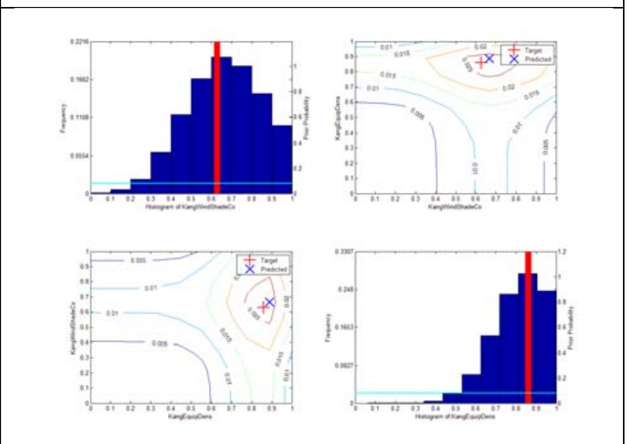
Discrete Variable setting Continuous Variable setting

Figure 4-15. Comparison of continuous and discrete parameter identification in DOE2.2 model

4.4.2.3.4. Sampling number

Sampling numbers were varied in this case study. Sampling numbers were set at 100, 1000, 10,000, 100,000 and the results are shown in Table 4-11. Except for the sampling number of 100, the remaining posterior distribution found the target value properly. The acceptance rates are similar for 1,000, 10,000, and 100,000 samples. As in the RC model case study, a sampling number of 1,000 is sufficient for the posterior distribution to find the target value. As a sampling number of 1,000 provides appropriate results, there is no need to have a higher sampling number to determine the posterior distribution.

Table 4-11. Posterior result with different sampling numbers in DOE2.2 model

	
100 sampling number CPU time: 5.2 seconds, Acceptance rate: 39%	1,000 sampling number CPU time: 34.9 seconds Acceptance rate: 42.8%
	
10,000 sampling number CPU time: 329.2 seconds Acceptance rate: 44.1%	100,000 sampling number CPU time: 3307.2 seconds Acceptance rate: 43.2%

4.4.3. *Reference building model*

4.4.3.1. *Description of the case study*

Several case studies were done in order to verify the validity of the parameter identification. The original case was derived from a commercial reference building (Office of Energy Efficiency & Renewable Energy). The U.S. Department of Energy (DOE) developed commercial reference building models, formerly known as commercial building benchmark models, in conjunction with three of its national laboratories. These are EnergyPlus models of 16 building types that represent approximately 70% of the commercial buildings in the U.S. Among those 16 building types, this case study chose a small office building that was constructed in or after 1980 (Office of Energy Efficiency & Renewable Energy). Its exterior surface is shown in Figure 4-16. To further accommodate this model with DOE2.2, this EnergyPlus model was converted manually to a DOE2.2 model and the results are shown in Figure 4-17. This DOE2.2 model mimicked EnergyPlus values by including schedule, interior exterior construction, internal gain, and an HVAC system that was described in the EnergyPlus model. The differences between EnergyPlus model and DOE2.2 model was about -12.79% for electricity and 17.66% for gas energy consumption.

Table 4-12. Other Input setting for DOE2.2 reference building modeling

Parameter name	Value
Total Floor Area (m2)	511
Floor to Ceiling Height	3.05
Available Fuel Types	Gas, Electricity
Building shape	Rectangle
Aspect Ratio	1.5
WWR	21.2%
R value for the ext wall (m2K/W)	1.09
R value for the roof (m2K/W)	2.37
U value for the window (W/m2K)	3.53
Window SHGC	0.41

Vent Rate (L/s/person)	10
Light loads (W/m2)	19.48
Equipment loads (W/m2)	8.07
Total number of people	28
Cooling set temp (F)	75.2
Heating set temp (F)	69.8
Cooling COP	3.39
Heating efficiency (%)	80
Location	Boulder, CO

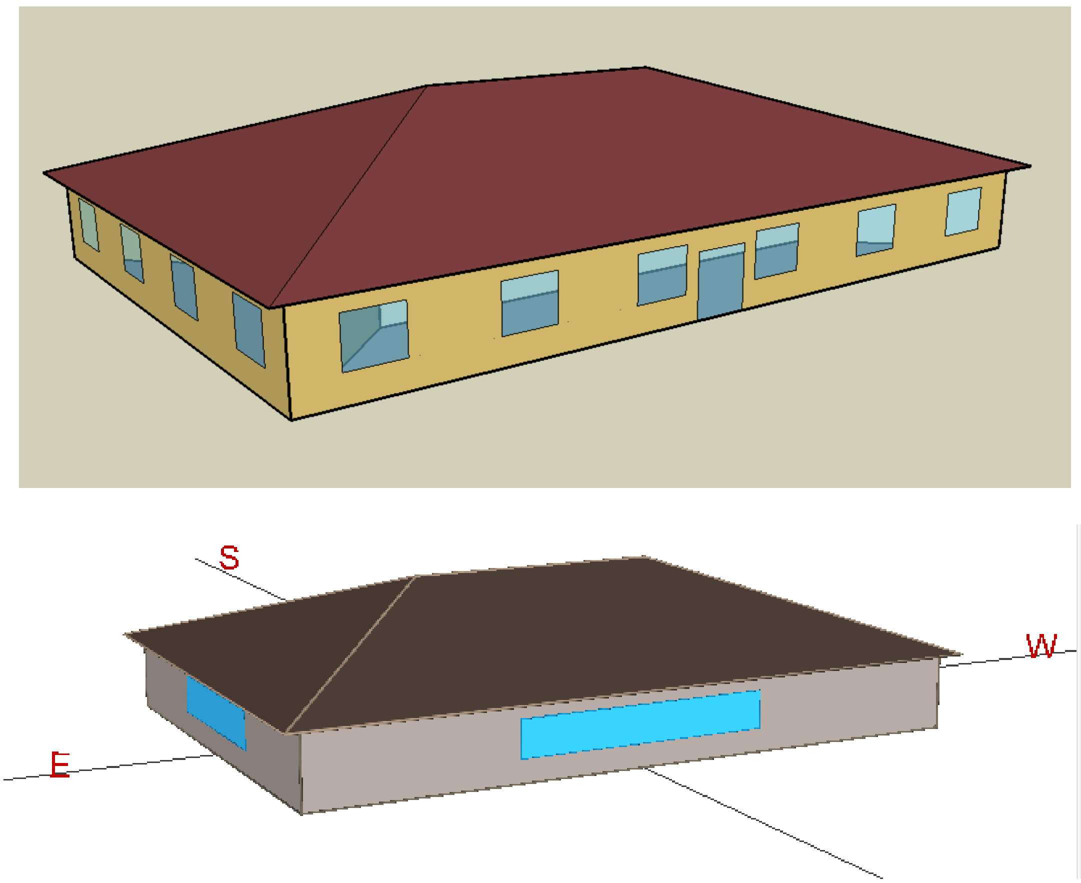


Figure 4-16. Appearance of exterior surface for EnergyPlus (up) and DOE2.2 model (down)

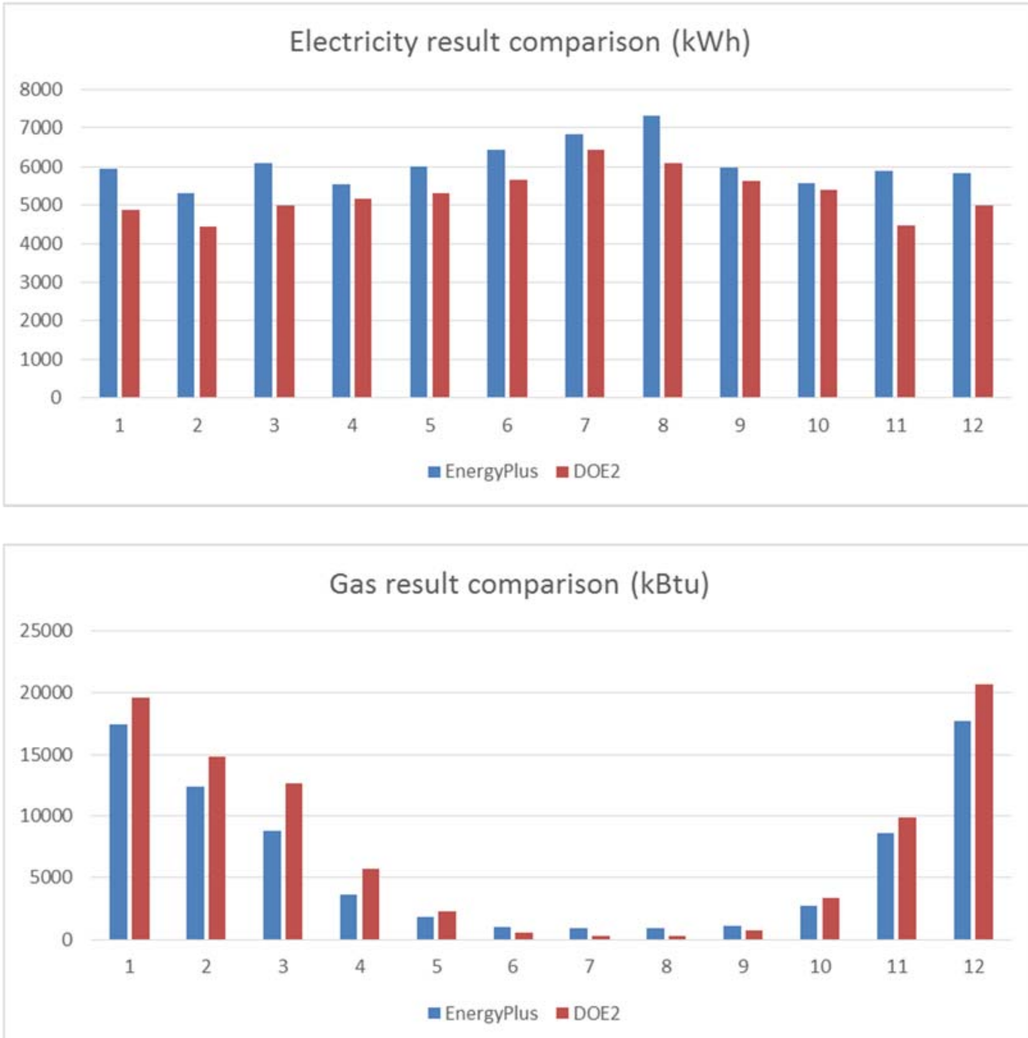


Figure 4-17. Comparison result between EnergyPlus and DOE2.2

4.4.3.2. Sensitivity analysis

As described in simple DOE2.2 model case study, this study will evaluate the reference building model's parameters. This study chose 11 variables to be varied within a Bayesian framework as shown in Table 4-13. This sensitivity result was calculated based on the default value (reference building's data) of each parameter and electricity and gas energy consumption results with a 10% variation for each. The results showed that variation in the heating set-point temperature had the largest energy consumption difference. Compared to the results from other variables, the impact on the energy consumption increased accordingly to a 10% variation. Figure 4-18 presents the results compared to the base case (reference building).

Table 4-13. Mean, Min, Max value and Sensitivity analysis result of input parameters for DOE2.2 reference bldg model

Description	Default Value	Min Value	Max Value	Rank	Impact on Energy Consumption difference (%) Per ± 10% variance	Impact on Energy Consumption difference (%) Per Min&Max variance
Roof Resistance (h-ft2-F/Btu)	17.63	2	32	11	0.04%	0.73%
Wall Resistance (h-ft2-F/Btu)	4.77	2	32	7	0.58%	11.67%
Window Shade Coefficient	0.39	0.3	1	10	0.09%	4.02%
Infiltration (CFM/ft2)	0.0474	0.01	2.01	8	0.49%	146.31%
Occupancy Density (ft2/person)	200	140	260	6	0.84%	9.13%
Light Density (W/ft2)	1.8098	0.3	3.3	2	4.10%	39.30%
Equipment Density (W/ft2)	1	0.3	3.3	3	2.83%	68.25%
Energy Input Ratio of Cooling	0.273	0.2	0.8	9	0.34%	14.15%
Heat Input Ratio of Heating	1.25	1	1.6	5	1.14%	13.63%
Cooling set temp (F)	75.2	60	80	4	2.12%	8.35%
Heating set temp (F)	69.8	60	80	1	4.33%	31.38%

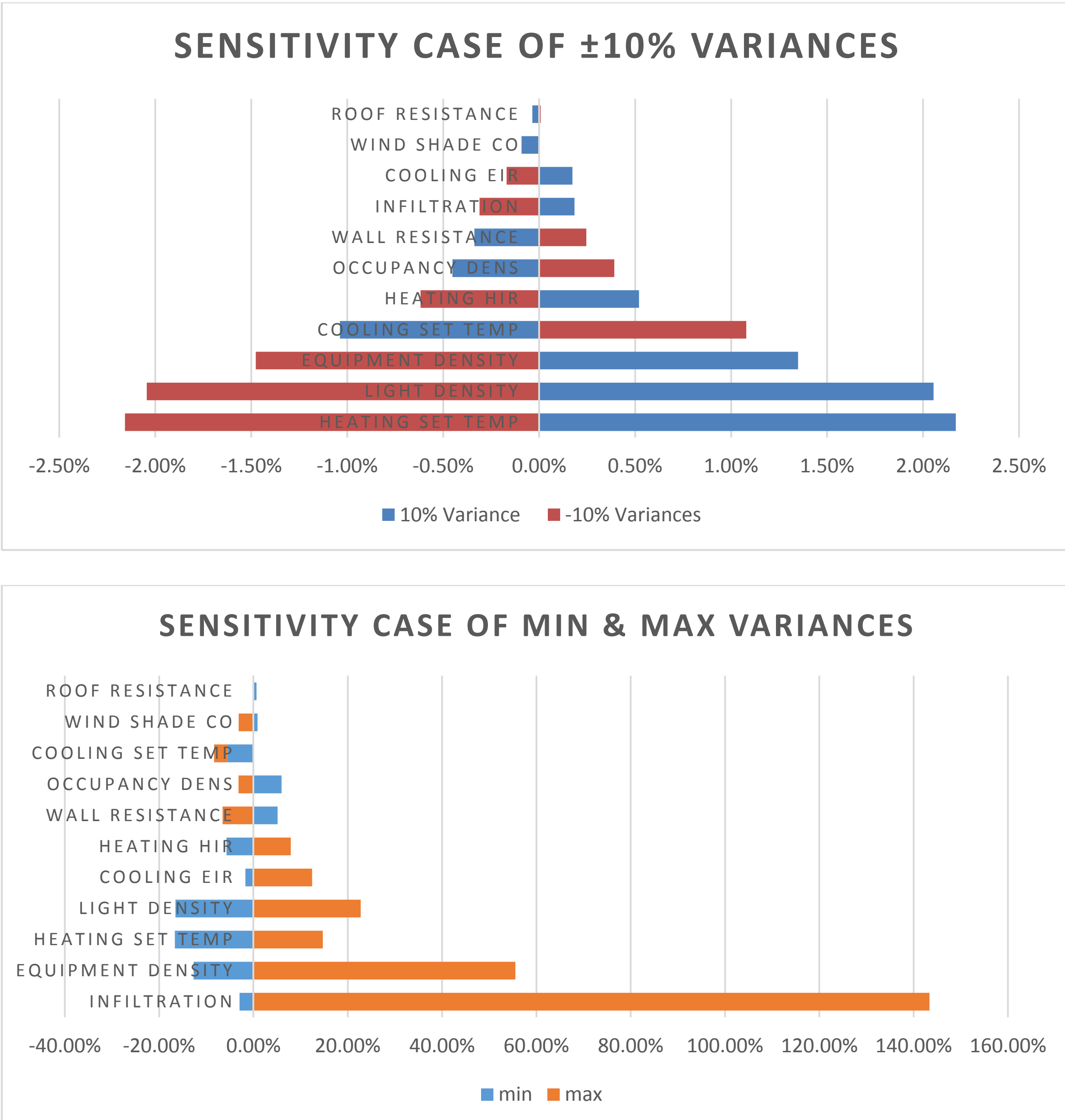


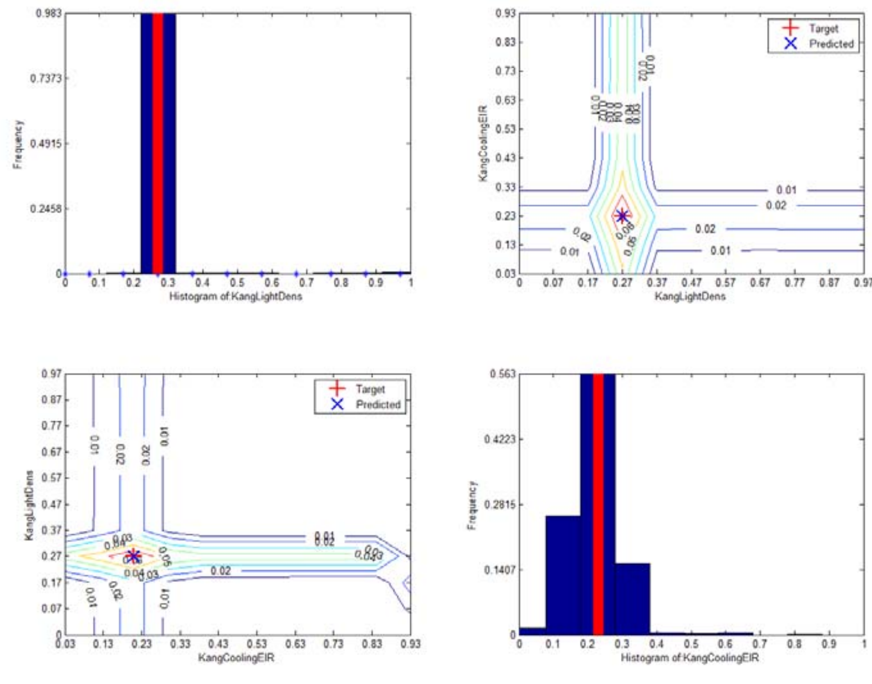
Figure 4-18. Sensitivity result of selected parameters for DOE2.2 reference bldg model

4.4.3.2.1. Observation noise

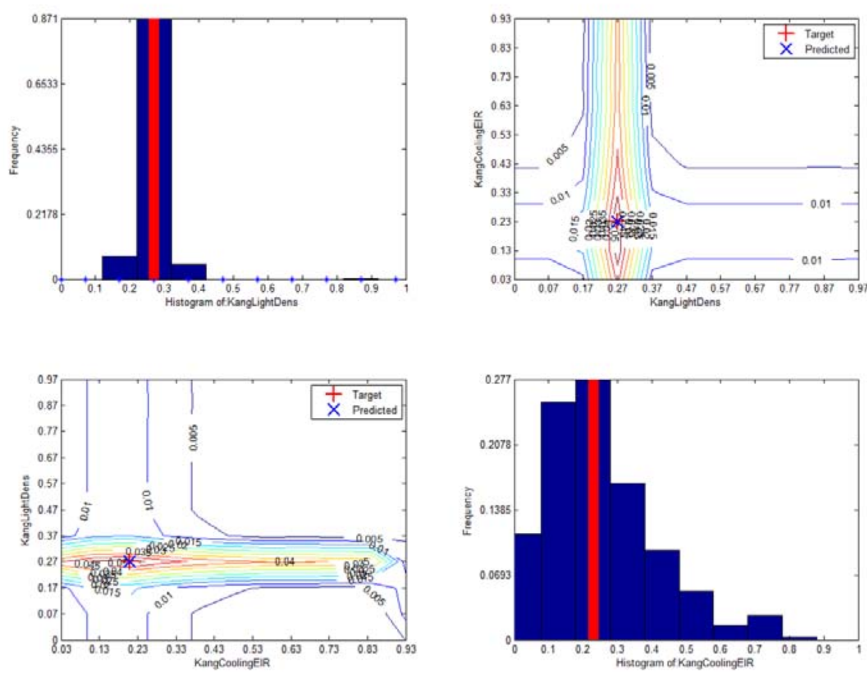
In this case study, sensitivity analysis was done for several values of observation noise. Higdon (Higdon, et al., 2004) stated that observation noise is the noise value from observing field measurements. There are two ways to obtain this value. One method is to set the observation error as another parameter that needs to be identified as with the others. Another way is to perform a sensitivity analysis on the observation noise and set a value beforehand. Table 4-14 presents variations in results for this. As with the sensitivity result in the RC model, a red bar line in the diagonal plot within each figure represents the target reference value that we are looking for and the graph itself is the posterior distribution after MCMC sampling. Lastly, the contour graphs depict the posterior result for each parameter when paired with other parameters. A plus marker '+' represents the predicted target value from a Bayesian framework and an X marker 'X' represents the original target value. The sampling number used in this case was 1,000. As shown in the table, all of the observation noise seems to identify a correct reference value. The acceptance rate of each case study tended to increase along with an increase in observation error.

For five unknown parameter cases, this study applied a prior distribution around the target value since the Bayesian framework kept failing to find a target value when applied with a uniform prior distribution. As shown in the table, if the observation error is more or less than $\sigma_y = e^{-10.5}$ (for two parameter case) and $\sigma_y = e^{-12}$ (for five parameter case), despite pointing to the right target value, the acceptance rate tends to increase or decrease following an observation increase or decrease. Since the acceptance rate should fit within 25% ~ 45%, an observation error of $\sigma_y = e^{-10.5}$ (for two parameter case) and $\sigma_y = e^{-12}$ (for five parameter case) should be chosen.

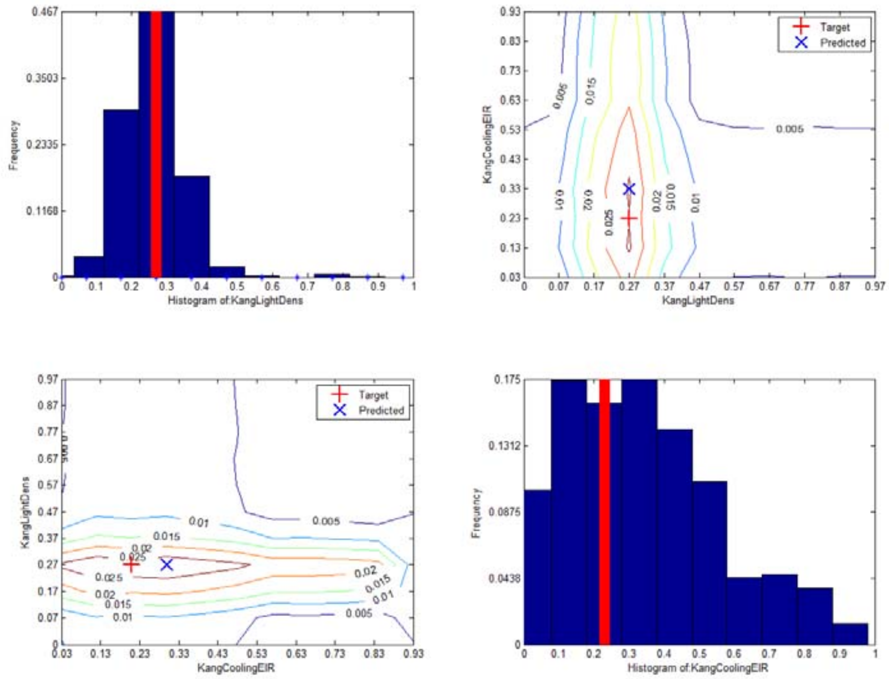
Table 4-14. Sensitivity result of the observation noise in reference DOE2.2 building model



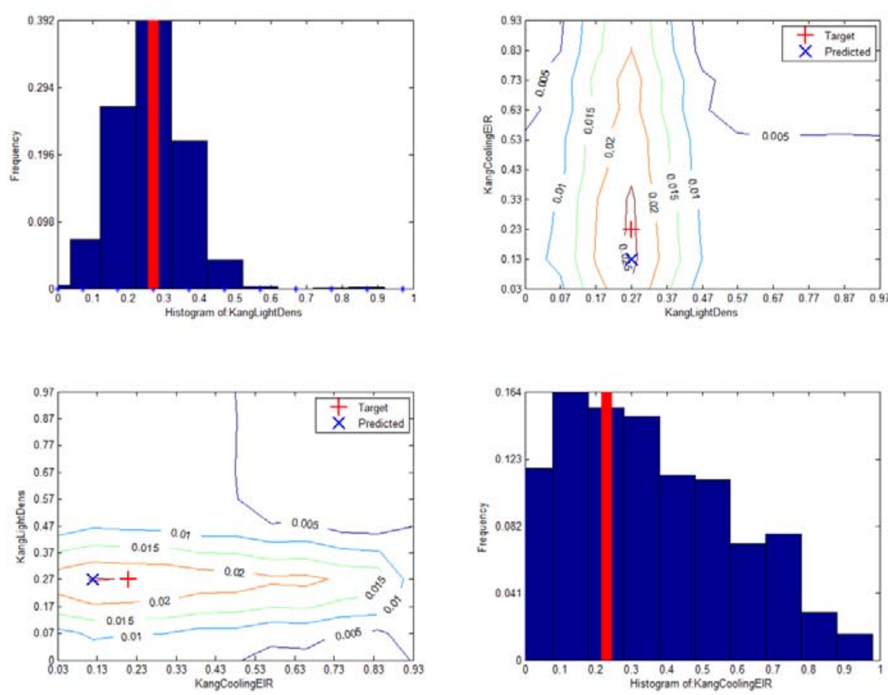
$\sigma_y = e^{-11.5}$ with 2 unknown parameters - CPU Time: 34.7 sec, acceptance rate: 20.2 %



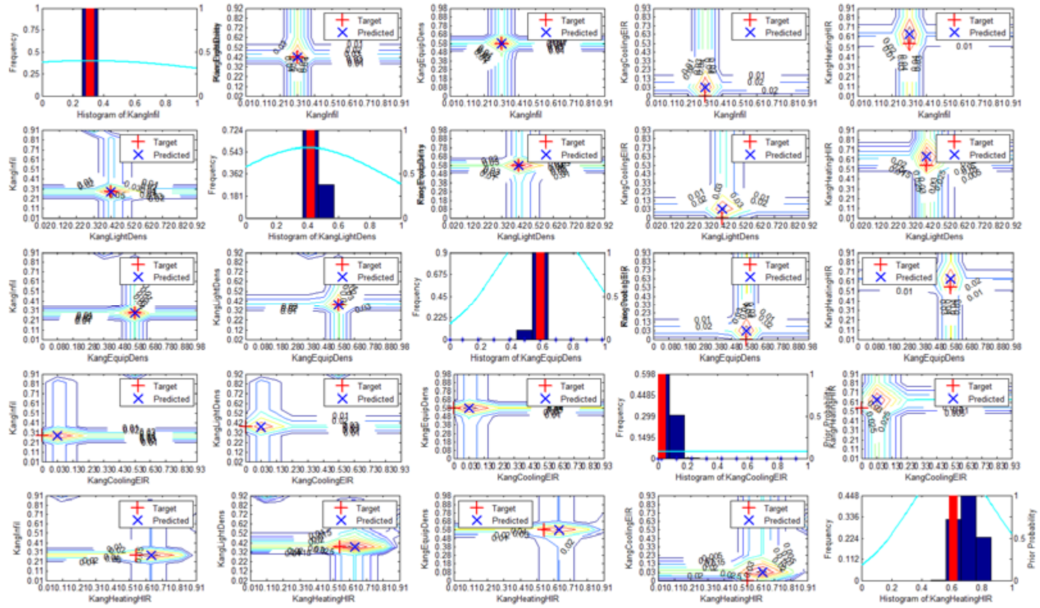
$\sigma_y = e^{-10.5}$ with 2 unknown parameters - CPU Time: 31 sec, acceptance rate: 30.6 %



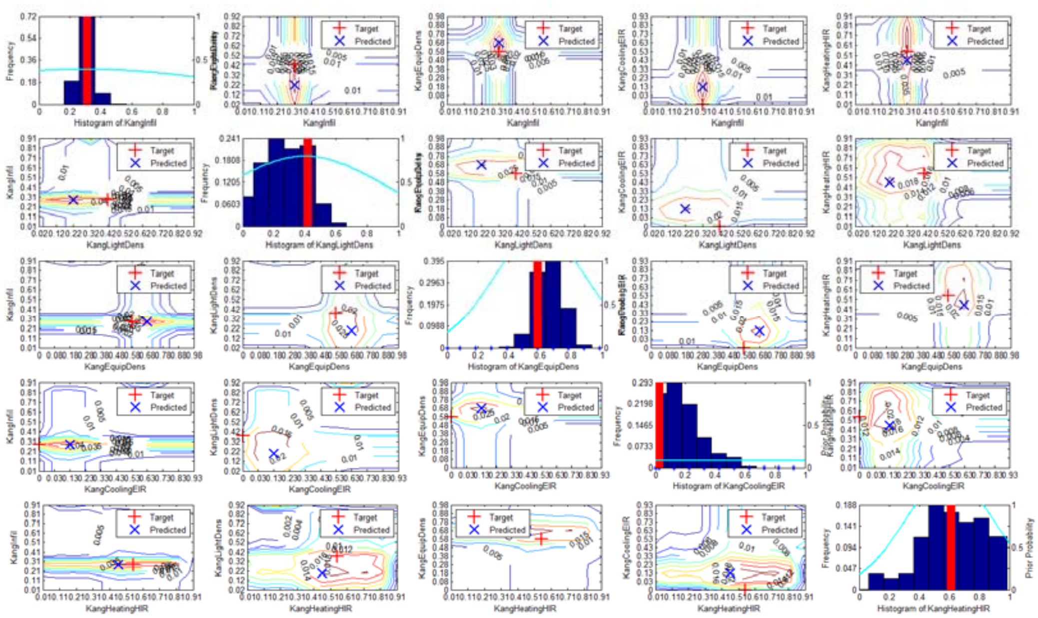
$\sigma_y = e^{-9.5}$ with 2 unknown parameters - CPU Time: 31.7 sec, acceptance rate: 56.7 %



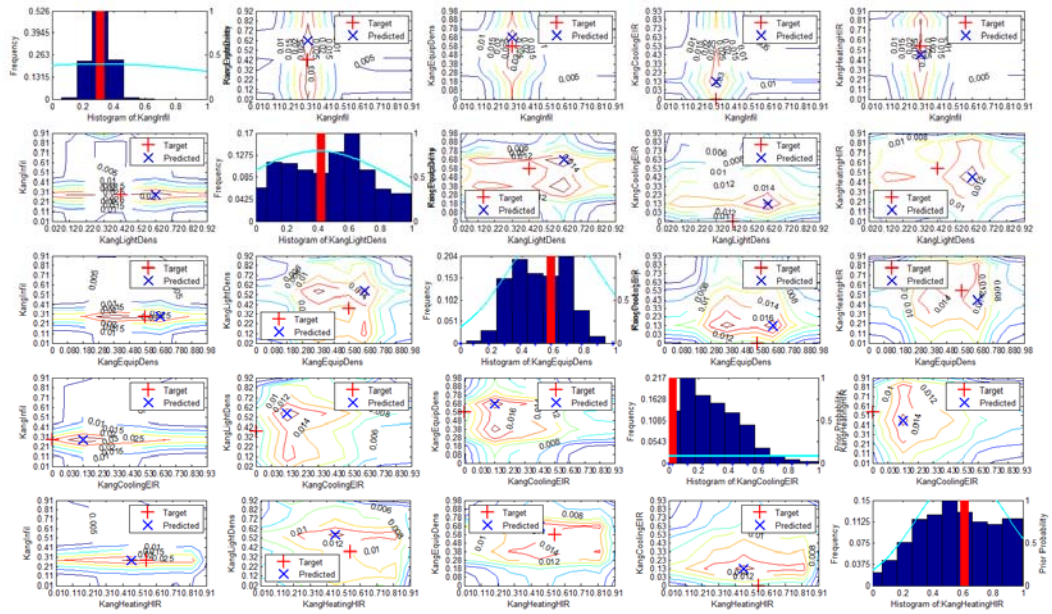
$\sigma_y = e^{-9}$ with 2 unknown parameters - CPU Time: 29.9 sec, acceptance rate: 62.4 %



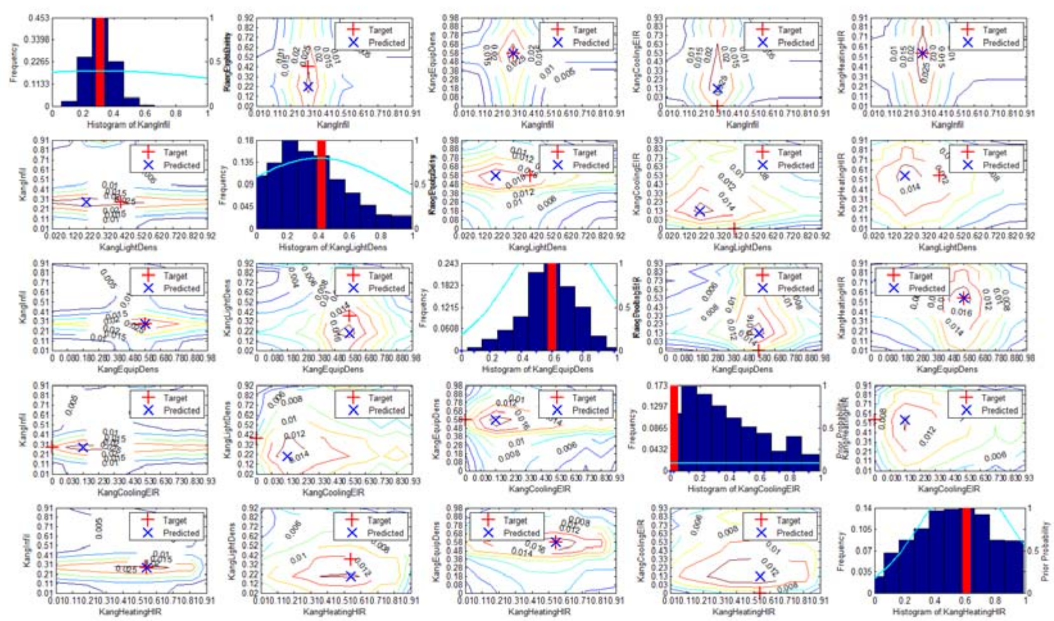
$\sigma_y = e^{-14.5}$ with 5 unknown parameters - CPU Time: 103.4 sec, acceptance rate: 10.8%



$\sigma_y = e^{-12}$ with 5 unknown parameters - CPU Time: 100.3 sec, acceptance rate: 30.5%



$\sigma_y = e^{-10.5}$ with 5 unknown parameters - CPU Time: 100.7 sec, acceptance rate: 48.9%



$\sigma_y = e^{-10}$ with 5 unknown parameters - CPU Time: 97.5 sec, acceptance rate: 53.2%

4.4.3.2.2. Size of Gaussian process

In Figure 4-19 (top), the average accuracy percentage error is shown in a plot. As suspected, a size of 100 had the most errors among the various sizes. These percentage errors were calculated based on 10,000 testing points for each size. In the eleven unknown model, the percentage error reaches 58.15%, which makes it useless as an emulator. With a size of 100, however, the average percentage error for 500, 1,000, 5,000 models were 2.3%, 1.93%, and 1.6%, respectively. To account for errors above 10%, Figure 4-19 (below) shows the number of values that exceeds 10% in errors for each size of a Gaussian process along with the computing time. A 10% threshold is important because in a normal situation, we don’t want the building model to have more than a 10% error criteria. Compared to the computing time, running an emulator with more than 5,000 in the training data seems useless since the building emulator itself takes more than seven hours and requires about 4.5 Gb to store the data. In the meantime, a size of 500 and 1,000 took less than an hour to build the Gaussian process and it had accurate results. The most accurate results came with 1,000 training points, especially in comparison to the results obtained with 500. Therefore, this study will utilize 1,000 as the size for the Gaussian process as an emulator.

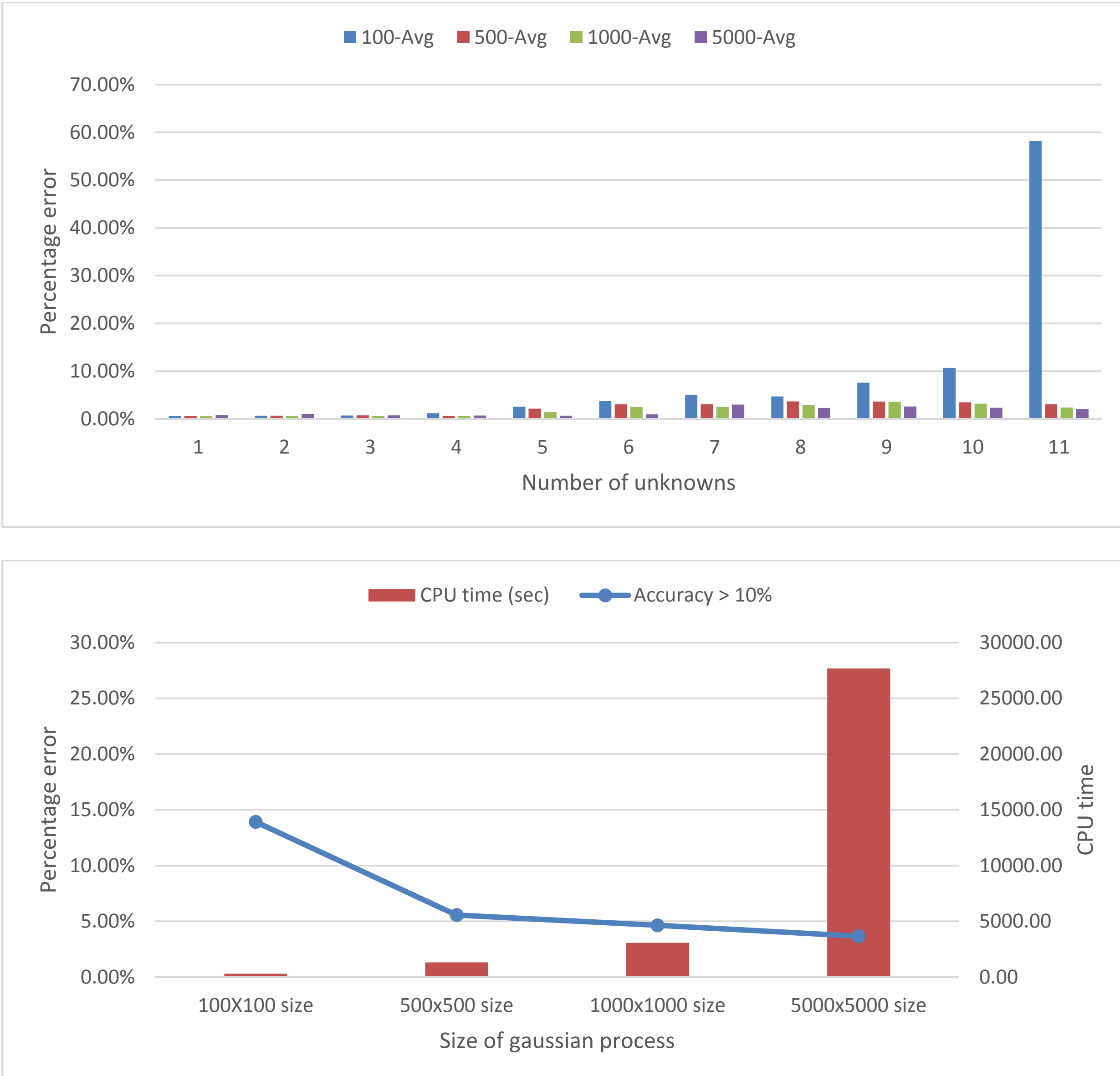
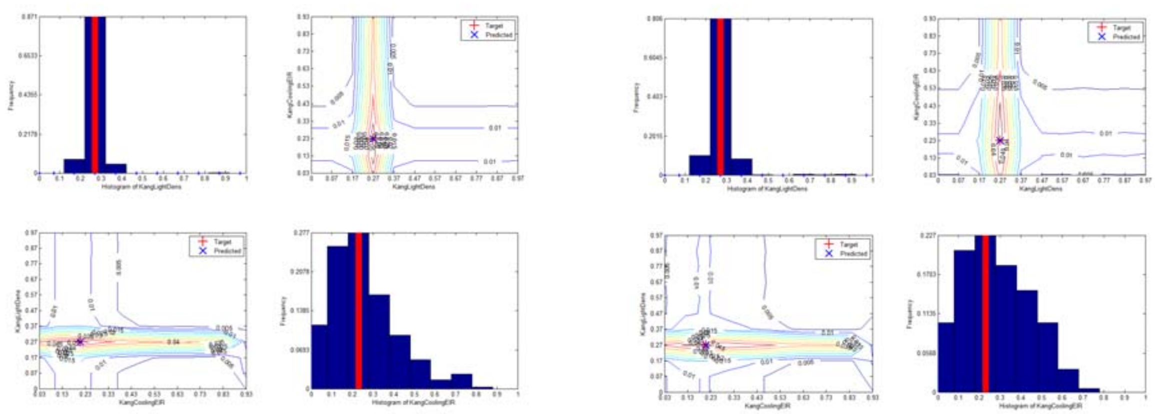


Figure 4-19. Size of Gaussian process results in DOE2.2 model

4.4.3.2.3. Continuous variables and discrete variables

In this study, a two unknown parameters identification case (lighting density and equipment density) for DOE2.2 with sampling number of 1,000 was performed with a discrete variable and continuous variable setting. To further explain, if a variable takes any value between two specified values, it can be called continuous, otherwise it is a discrete variable. For instance, glass type is representative parameter of a discrete variable. This case study evaluated the approach with both a continuous and discrete variable setting to determine the influence on accuracy as well as acceptance rate. In Figure 4-20, both cases seem to find the appropriate target value despite having the variable set at either 'Discrete' or 'Continuous'. In terms of accuracy, the discrete variable setting had a steeper result (more frequently sampled). This was because there weren't many fields to explore compared to the continuous variable setting. Small search field can lead to higher occurrences in poster distribution. The acceptance rate for each case were 27.2% (discrete) and 40.6% (continuous). These results suggest that the discrete variable setting refused more samples than the continuous variable setting.



Discrete Variable setting　　　　　　Continuous Variable setting

Figure 4-20. Comparison of continuous and discrete parameter identification in DOE2.2 model

4.4.3.2.4. Number of sampling

In this section, sampling numbers were varied to verify the impact on the posterior distribution. Sampling number is the number of sampling using Metroplis Hastings algorithms.

Generally, a larger number will have a more accurate result. Sampling numbers of 100, 1,000, 10,000, 100,000 were tested and the results are shown in Table 4-15. For a sampling number of 100, due to the fact that the sampling number is too small, the analysis failed to identify the appropriate target values. The rest of the posterior distribution finds the target value properly. Other cases identified the parameter appropriately. The acceptance rate seems to remain around 40%. This indicates that when the sampling number is above 1,000, it is sufficient to find the target value. In the table, the frequency of the target value refers to the posterior occurrences on each target value as indicated by a red bar in the histogram. There are two aspects that need to be considered for the number of sampling case study. The first is accuracy and the second is the computing demand. Figure 4-21 shows the results in a comparison computing time and acceptance rate. As shown in the figure, the acceptance rate is similar with sampling numbers above 100. However, computing time increases exponentially when the sampling number increases. In our case, a sampling number of 1,000 was selected because it provides similar results as the higher sampling numbers and requires only a very short amount of time to do the whole sampling.

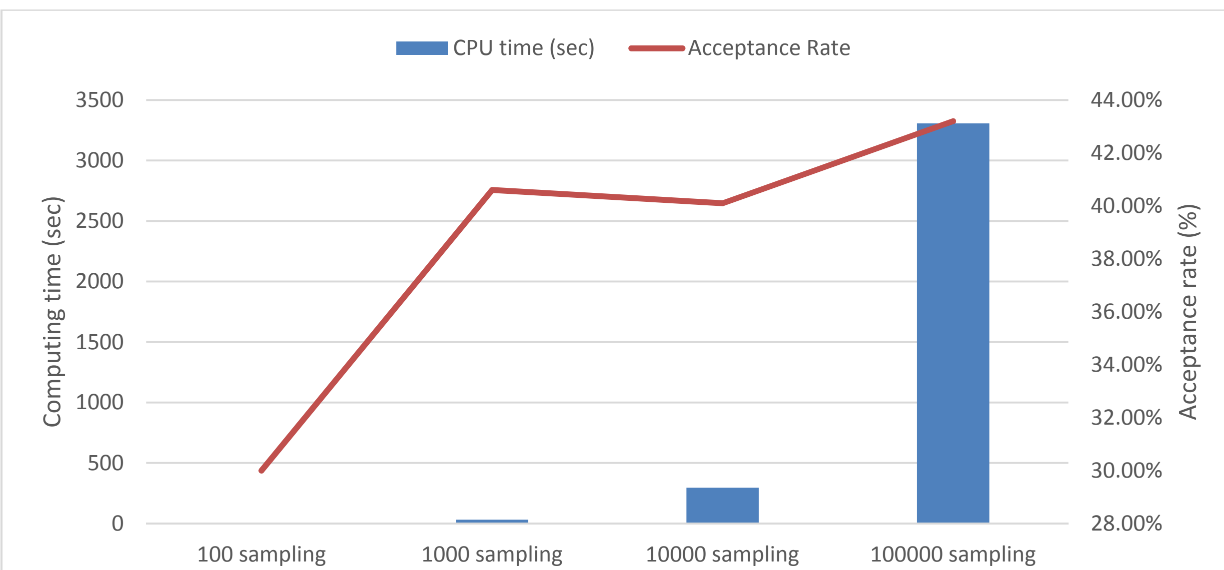
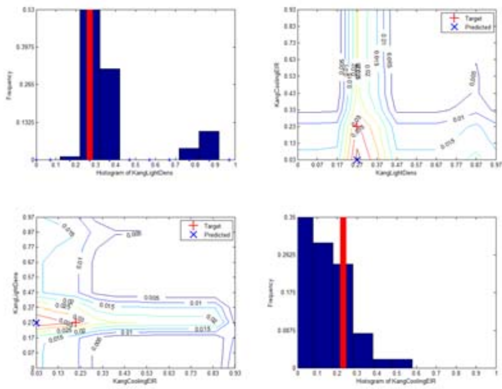
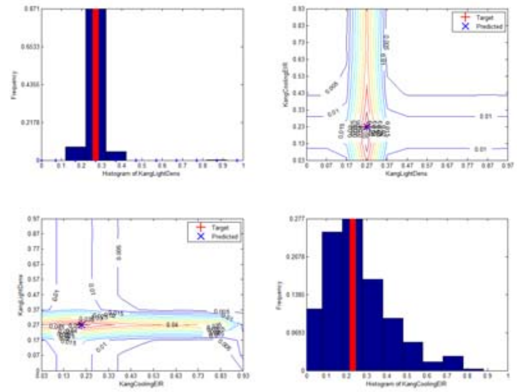
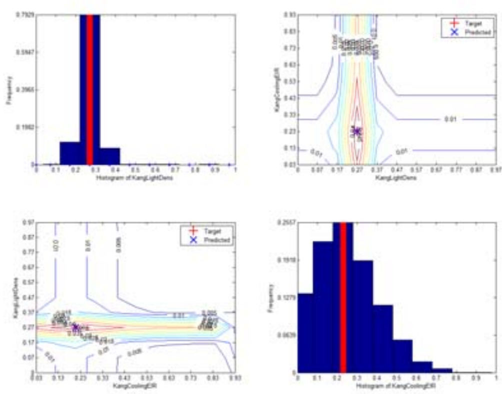
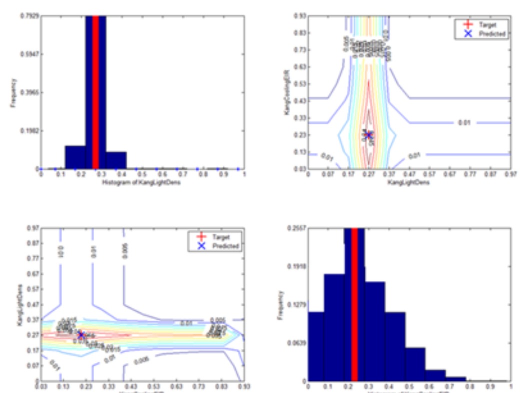


Figure 4-21. Sampling number variance result for DOE2.2

Table 4-15. Posterior result with different sampling numbers in DOE2.2 model

	
100 sampling number CPU time: 4.6 seconds, Acceptance rate: 30% Frequency of target value: 38%	1,000 sampling number CPU time: 31.0 seconds Acceptance rate: 40.6% Frequency of target value: 50.5%
	
10,000 sampling number CPU time: 295.2 seconds Acceptance rate: 40.1% Frequency of target value: 52.5%	100,000 sampling number CPU time: 3307.2 seconds Acceptance rate: 43.2% Frequency of target value: 53%

4.4.3.2.5. Sensitivity analysis on number of unknowns

In the previous sections, this study performed several case studies to determine validity on the observation noise, size of the Gaussian process, a continuous/discrete setting, and number of sampling. In this section, this study performs sensitivity analysis for a number of unknowns. The suspected number of unknowns are critical to parameter identifications. A higher number of unknowns creates more uncertainty and results in a difficult situation for parameter identification. This study increased the number of unknowns to verify capability for parameter identification. To validate the case study results, this study used Posterior Root Mean Square Error (PRMSE) for a comparison of the results as shown in Eq. (4.8).

$$WA_i = \frac{\sum_{j=1}^{m}(O_j * \mu_j)}{Sampling\ Number} \tag{4.8}$$

$$Posterior\ Root\ Mean\ Square\ Error\ (PRMSE) = \frac{\sqrt{\sum_{i=1}^{n}(WA_i - Target_i)^2}}{n}$$

Where, j = m number of bins for histogram occurrences, i = n number of unknowns, O_j is the number of occurrences and μ_j is the mean of each bin in a histogram. WA_i is the ith weighted average calculated. Lastly, posterior error is determined by calculating the percentage error of the weighted average and target value and divided by the number of unknowns that have an average posterior error as shown in Eq. (4.8).

As shown in Figure 4-22, CPU time and posterior error increased gradually as the number of unknowns increased. Despite the number of unknowns reaching the maximum (11 unknowns), the average posterior error stayed within 20%, which in this case can be interpreted as ‘fair’. If the number of unknowns were within 5, the posterior stayed within a 10% error. Additionally, with 11 unknowns, the cpu time only took about 210 seconds to sample the posterior distribution. This is acceptable for practical use of this framework.

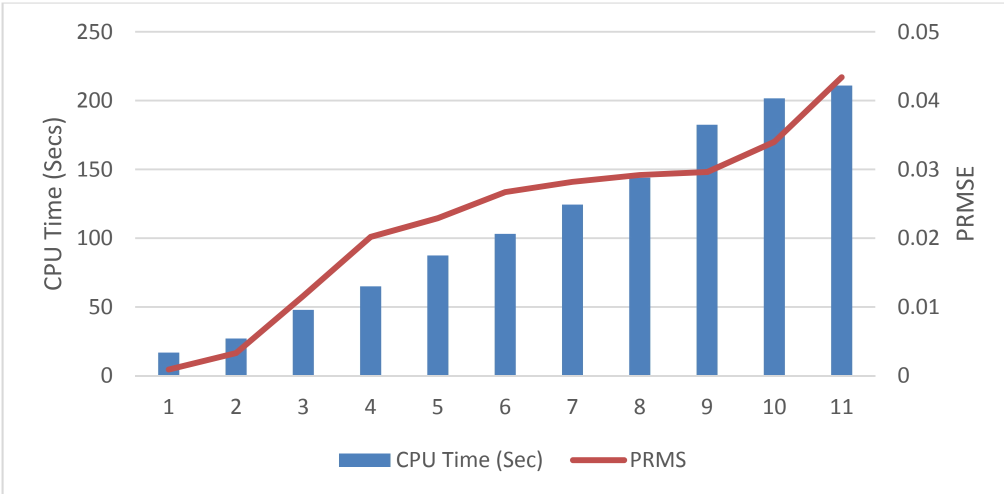


Figure 4-22. Sensitivity result for number of unknowns in DOE2.2 reference bldg. model

4.5. Summary and Conclusions

Until now, most parameter identification problems were solved by optimization methods. However, this study proposed a statistical and probabilistic way to approach this type of issue. A Bayesian approach was proposed to perform parameter identification. Compared to a traditional Bayesian approach, this approach eliminated the bottleneck process within a traditional Bayesian approach and applied an emulator to replace the simulation data. With this process, this study has several sensitivity case studies to verify validity.

The sensitivity case studies suggested that by using various observation noise values depending on each energy model cases with a 1,000 sampling number provided the optimal results for this approach. Table 4-16 suggests the observation result for each energy model cases. A Gaussian process size of 1,000 was sufficient to act as an emulator and this approach showed potential for use with either discrete or continuous variables. The number of unknown parameters showed that this approach can applied in a situation with multiple unknowns. This is, however, limited as an increase in uncertainty coincides with an increase in the number of unknowns. To eliminate these uncertainties, this study suggested using higher prior probability unknowns.

Table 4-16. The observation noise result for various energy model cases

	2 unknown parameter case	**5 unknown parameter case**
RC office model	$\sigma_y = e^{-6.5}$	$\sigma_y = e^{-4}$
DOE2.2 Simple model	$\sigma_y = e^{-7.5}$	$\sigma_y = e^{-8}$
DOE2.2 Reference bldg. model	$\sigma_y = e^{-10.5}$	$\sigma_y = e^{-12}$

In order to apply our work to practical use, further research is required in several areas. For example, it would be necessary to test a methodology in actual buildings instead of reference buildings. There may be more uncertainties involved in actual situations as reference buildings only mimic the typical building behaviors. Additionally, the reference buildings were created using a one-zone system, while actual buildings typically utilize multi-zoning schemes. To test the methodology in practice, validation is necessary in several conditions.

This approach was only tested for an unknown parameter identification case. There are more applications that can utilize this approach, such as an energy audit. In order to do that, this approach must be tested under several applications. Validation with energy conservation measure identification would be one example.

Our approach also needs to be compared to other optimizations. Given there are numerous optimization methods available and with different advantages and disadvantages. This study requires further verification of our findings in comparison to other approaches in order to validate the capabilities of this methodology before this approach can be utilized in practice.

CHAPTER 5: Case study of unknown parameter identification

5.1. Introduction

Until now, various sensitivity analyses were performed to verify the validity for the unknown parameter identification. This study has found that the proposed approach can be utilized effectively in identifying unknown parameters in closed loop cases. In this section, this study will perform case studies of actual existing buildings. In reality, building an energy model for a building requires both expertise and effort from the modeler. Many prior studies tried several approaches to alleviate these burden. For example, various optimization methodologies for unknown parameter identification were identified in the literature review. However, these approaches had a blind spot in terms of targeting its main objective as matching the reference data only. This can lead to erroneous parameter identification due to the characteristic of matching the reference data deterministically. In comparison to other deterministic optimization approaches, the Bayesian approach utilizes probability and statistics to determine the unknown parameters. This approach has the benefit of identifying the parameter with more flexibility. However, a traditional Bayesian approach has a downside in consuming considerable computing time. In this study, we have proposed a Bayesian approach with a meta-model (emulator) to eliminate the bottleneck procedure that exists in the traditional Bayesian approach. The proposed approach succeeded in identifying unknown parameters and was verified in several sensitivity case scenarios.

This section validates the proposed approach with an existing building in order to determine its practicality.

5.2. Case study of medium office building

5.2.1. Description of case study

The first building was one chosen from Boulder, Colorado. The building is a multi-use two-story commercial building with a floor area of roughly 17,000 ft2. The second level of the building is primarily office space, although there are also three conference rooms, a small data closet, a reception area, and two kitchens. The ground floor of the building is dominated by a large storage area and a construction room, where workers repair appliances, doors, and complete other small apartment-related projects. The remainder of the ground floor is comprised of another kitchen, conference room, office space, storage, and a garage for grounds-keeping equipment. Part of the second floor is above a small covered parking area and an outdoor stairwell and entry area. It is a well-constructed building, however, the building has a high level of energy consumption due to the following reasons: a complex multi-zone HVAC system with several localized user-manipulated thermostats, an offsite set-point manager, and multiple operable windows, etc. This makes the building's energy use profile rather complex for a building energy model, and therefore has uncertainty in its energy model. Figure 5-1 depicts the exterior, its blueprint and its eQUEST modeling screen.

Figure 5-1. Exterior (top-left) of medium office building and its blueprint (top-right) and its eQUEST modeling screen (bottom)

5.2.2. Selection of unknown parameters

To be used in the proposed approach, this study has chosen the following variables for the identification process. These parameters were chosen because of the following reasons: the values can be easily obtained from the blueprint, these values contains higher uncertainty, and these values will be manipulated prior to other parameters in energy modeling procedures. Table 5-1 shows these parameters and their values. Within the table, the values without any underline suggest the actual values that we obtained from a field audit or from the blueprint. In contrast, values with an underline are values that are unknown and those we wish to identify. In the minimum and maximum value section, the limit of the each parameter can be set by users. For the initial run, unknown values were set to the default value that was obtained from eQUEST simulations. For example, if the default value for the underlined parameter is set to 1, then this value is not an exact value from the source. Rather, this value was found from eQUEST simulation. Since there are several zones with different densities each for a certain parameter, this study has decided to use a multiplier instead of using actual values. By '1', it means that the value for each zone will be multiplied by 1.

Table 5-1. Parameters for medium office building (default, minimum & maximum value setting)

Description	Default Value	Min Value	Max Value	Rank	Impact on Energy Consumption difference (%) Per ± 10% variance	Note
Roof U Value (Btu/h-ft2-F)	0.036	0.01	0.5	8	0.18%	
Wall U Value (Btu/h-ft2-F)	0.043	0.01	0.5	10	0.1679%	
<u>Infiltration (ACH)</u>	**<u>0.5</u>**	**<u>0.01</u>**	**<u>2.01</u>**	**<u>3</u>**	**<u>2.65%</u>**	
Occupancy Density (ft2/person)	250	100	400	5	0.87%	
Light Density (W/ft2)	1	0.01	1.5	6	0.75%	This value represents multiplier
<u>Equipment Density (W/ft2)</u>	**<u>1.285</u>**	**<u>0.01</u>**	**<u>1.5</u>**	**<u>2</u>**	**<u>2.82%</u>**	
Energy Input Ratio of Cooling	0.29	0.01	0.8	7	0.69%	
<u>Heat Input Ratio of Heating</u>	**<u>1.35</u>**	**<u>1</u>**	**<u>1.6</u>**	**<u>1</u>**	**<u>2.97%</u>**	
Cooling set temp (F)	75	60	80	9	0.1688%	
<u>Heating set temp (F)</u>	**<u>72</u>**	**<u>60</u>**	**<u>80</u>**	**<u>4</u>**	**<u>0.91%</u>**	

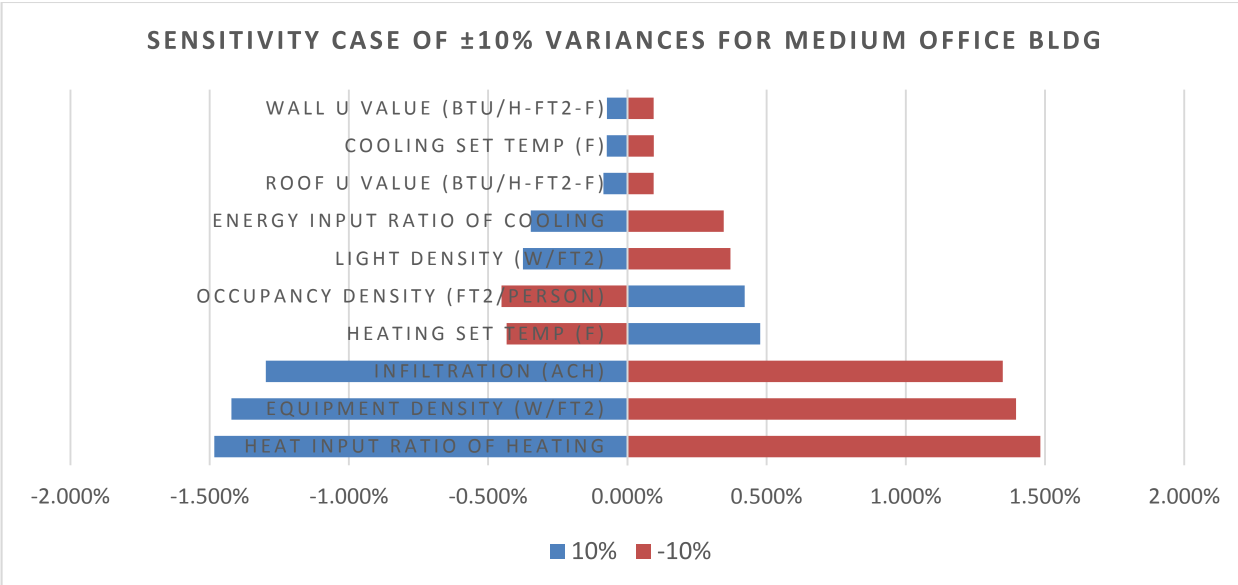


Figure 5-2. Sensitivity result of medium office building after ±10% variations

After sensitivity analysis on the impact of each parameter on energy consumption, this study ranked each parameter for later use. Figure 5-2 presents the sensitivity results. The results show that the heat input ratio for heating has the highest impact on energy consumption. In contrast, the u-value of a wall has the least impact. By using values from the above table, an initial run was performed to verify the impact of these default values. As shown in Figure 5-3, there was a large difference in electricity and gas usage compared to utility data.

Figure 5-3. Initial run result for medium office building

5.2.3. Parameter identification for medium office building

Figure 5-4 shows the posterior distribution results for four unknown parameter identification of a medium office building case. The total cpu time for this case with the sampling number of 1,000 was about 93 seconds. The posterior distributions of infiltration, equipment density, heating input ratio, and heating set-point temperature are shown in the figure. All of the parameters had a peak occurrence at a certain location. By analyzing these posterior distributions, we can set theses peak occurrence locations for each parameter as predicted values from the Bayesian approach. Subsequently, we can re-run the original case with these values. Three run scenarios were created using the mean, minimum, maximum values for these predicted values. Figure 5-5 shows these results. Compared to the initial run, the subsequent run showed the energy consumption result aligned to utility data. In this case, the process of building an emulator took about 90 minutes with a 0.38 % average error.

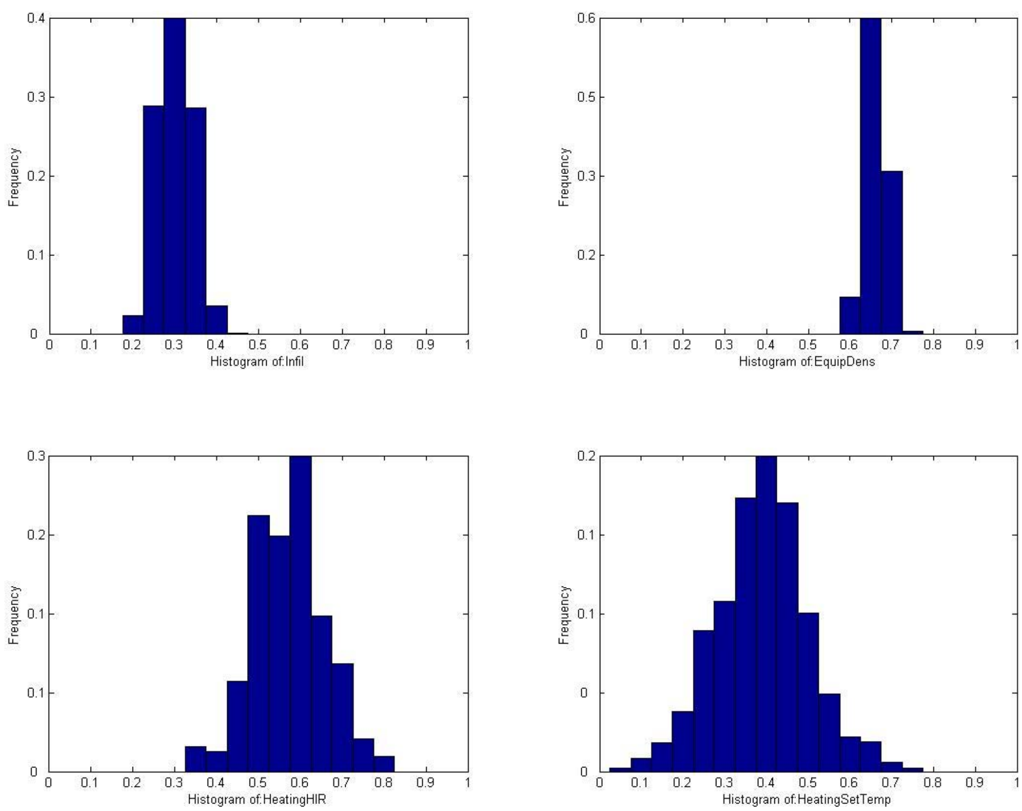
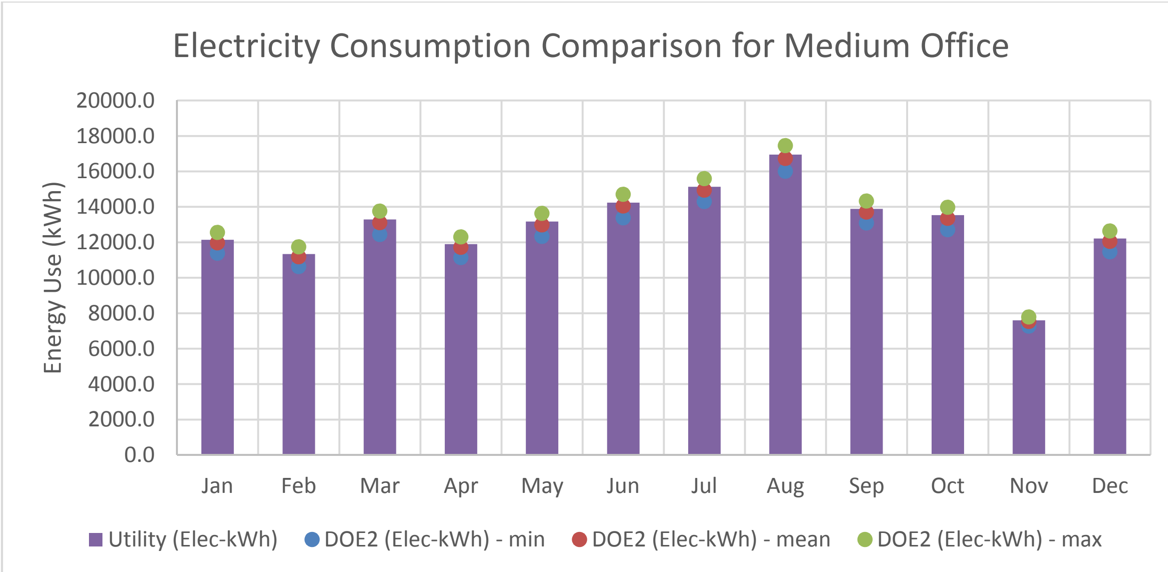


Figure 5-4. Unknown parameter identification result for medium office building



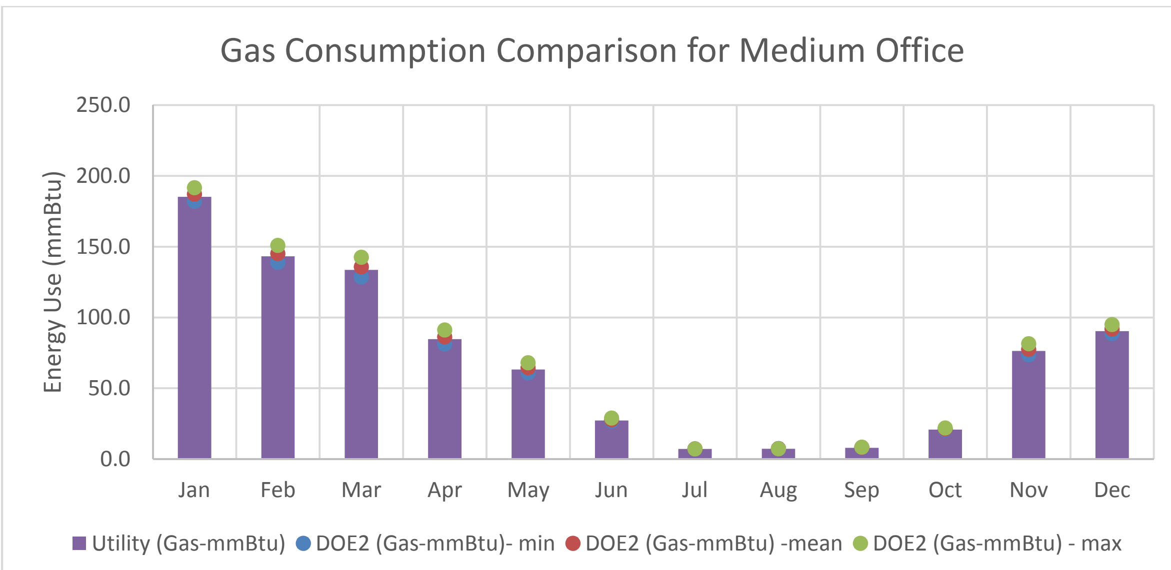
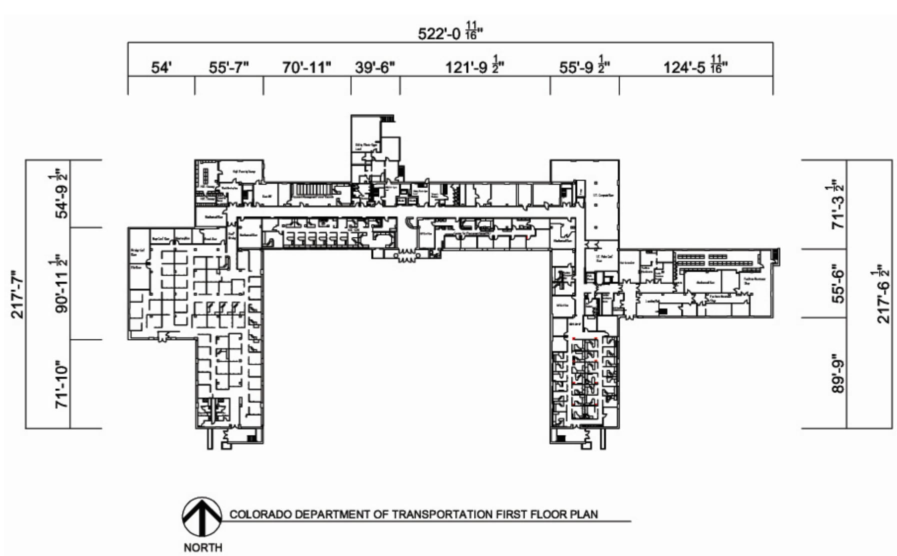


Figure 5-5. After run with using predicted values for medium office building

5.3. Case study of large office building

5.3.1. Description of case study

The second building chosen was located in Denver, CO. This building is a 4-story building with 203,135 sqft floor area. Office hours are 8 a.m. to 5 p.m., Monday to Friday. It is a southfacing building. Potentially, this building may utilize daylighting. However, since there are no tall objects around the building, which can block the direct sunlight, this building is exposed to direct sunlight during almost all office hours. Exposing to sunlight may affect building energy consumption and occupants directly or indirectly. The building is a “U” shaped building with the first two floors following the “U” shape, which creates wings that surround a central driveway loop. There are two additional floors above only the central section of the building. Figure 5-6 shows the exterior, a blueprint of the building, and the energy model.



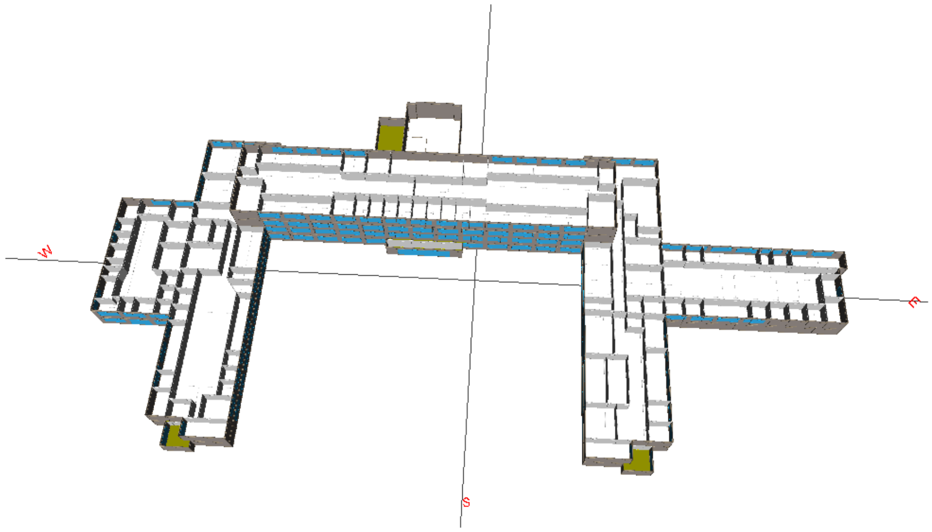


Figure 5-6. Exterior (top) of large office building and its blueprint (middle) and its eQUEST modeling screen (bottom)

There are twelve constant volume, multi-zone air handlers with steam heating and chilled water cooling. Typical air distribution diffusers are installed. The original intent of the HVAC design was to supply conditioned air to open office spaces. These spaces have since been framed in with minimal changes to the ducts or diffusers. This creates possible thermal comfort problems. Currently there is no return duct and limited flow through the ceiling plenum.

5.3.2. Selection of unknown parameters

To be used in the proposed approach, this study chose the following parameters for the identification process. Ten parameters were chosen in this case study and Table 5-2 shows these parameters. Similar to the medium office building case, the values without an underline suggests the actual calibrated value from a field audit or from the blueprint. In contrast, values with an underline suggest unknowns and were therefore obtained from eQuest default values. In the minimum and maximum value section, the limit of the each parameter can be set by users. Note that in the light and equipment density sections, all of the parameters were combined and generalized by dividing by its floor area. As shown in the table, the heat input ratio of heating was the most influential parameter.

Table 5-2. Parameters for large office building (default, minimum & maximum value setting)

Description	Default Value	Min Value	Max Value	Rank	Impact on Energy Consumption difference (%) Per ± 10% variance	Note
Roof U Value (Btu/h-ft2-F)	0.259	0.01	0.5	5	0.99%	
Wall U Value (Btu/h-ft2-F)	**0.08**	**0.01**	**0.5**	**4**	**1.03%**	
Window U Value (Btu/h-ft2-F)	5.82	1	10	6	0.87%	
Infiltration (ACH)	0.5	0.01	2.01	8	0.47%	
Light Density (W/ft2)	**1.257**	**0.01**	**1**	**2**	**1.58%**	**Generalized Value**
Equipment Density (W/ft2)	0.2	0.01	1	9	0.27%	Generalized Value
Energy Input Ratio of Cooling	0.1683	0.01	0.8	7	0.53%	
Heat Input Ratio of Heating	**1.25**	**1**	**1.6**	**1**	**1.90%**	
Cooling set temp (F)	**75**	**60**	**80**	**3**	**1.24%**	
Heating set temp (F)	68	60	80	10	0.18%	

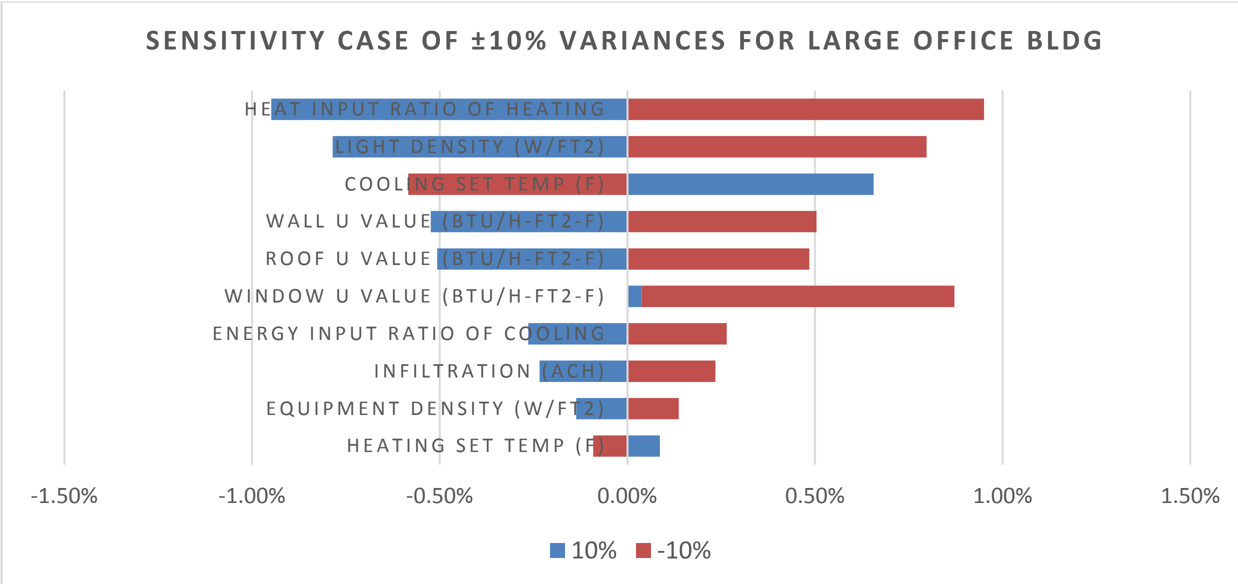


Figure 5-7. Sensitivity result of large office building after ±10% variations

After defining parameter limits and default values, a sensitivity case that varied by ±10% for its original case was performed. Figure 5-7 presents the sensitivity results. Strangely, the window u-value seems to increase when its value was increased by 10%, and on the contrary, energy consumption was increased when its value was decreased by 10%. This is due to the characteristic of a window that when it reaches a certain u-value, the energy usage will not increase gradually. Rather, it will increase despite a decrease of the value.

Figure 5-8 shows results for the initial run of large office building case. As shown in the figure, there was a large 26% monthly difference in electricity and 45% in gas. This suggests that the default values from eQUEST do not fit this model.

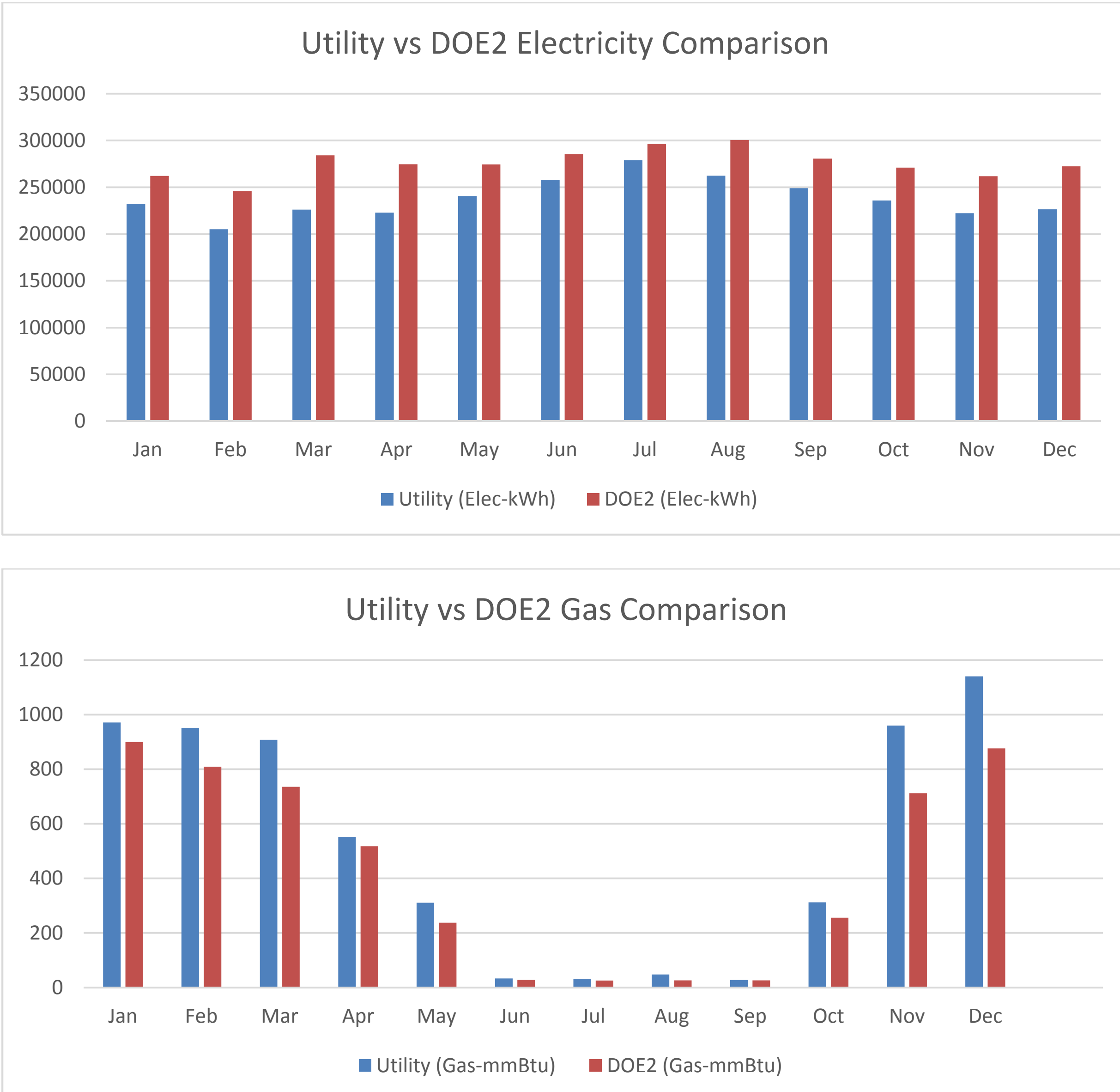


Figure 5-8. Initial run result for large office building

5.3.3. Parameter identification for Large office building

Following the sensitivity case results, this study chose a four unknown parameter identification case to verify validity. Figure 5-9 shows the parameter identification results. As shown in the figure, all four unknown parameters were identified. The selected parameters to be identified were: u-value of wall, lighting density, heating input ratio, and cooling set-point temperature. Similar to the medium office building, all of the parameters converged on some predicted values. By using these values, this study performed an after-run and Figure 5-10 shows the results. As shown in the figure, all of the parameters reached within 10% of the utility data. However, the total cpu time required to build the emulator was about 22 hours since the original DOE2.2 energy model took almost a minute to run a single simulation. This needs to be considered since total cpu time is crucial to achieving practical parameter identification.

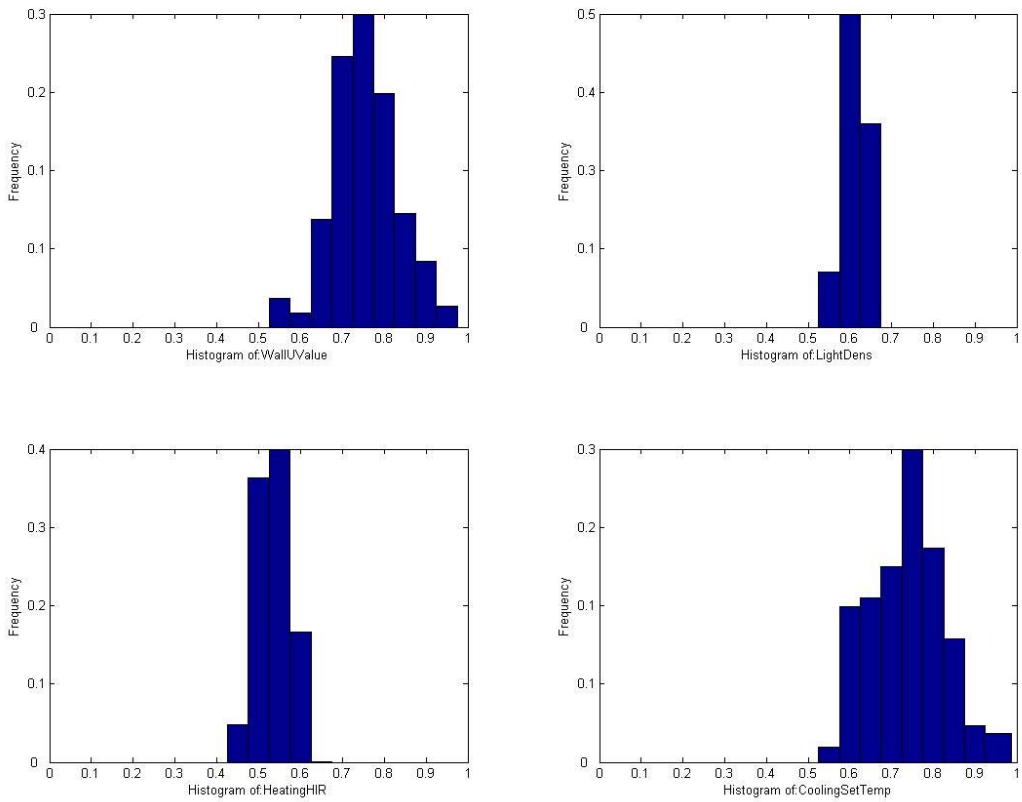
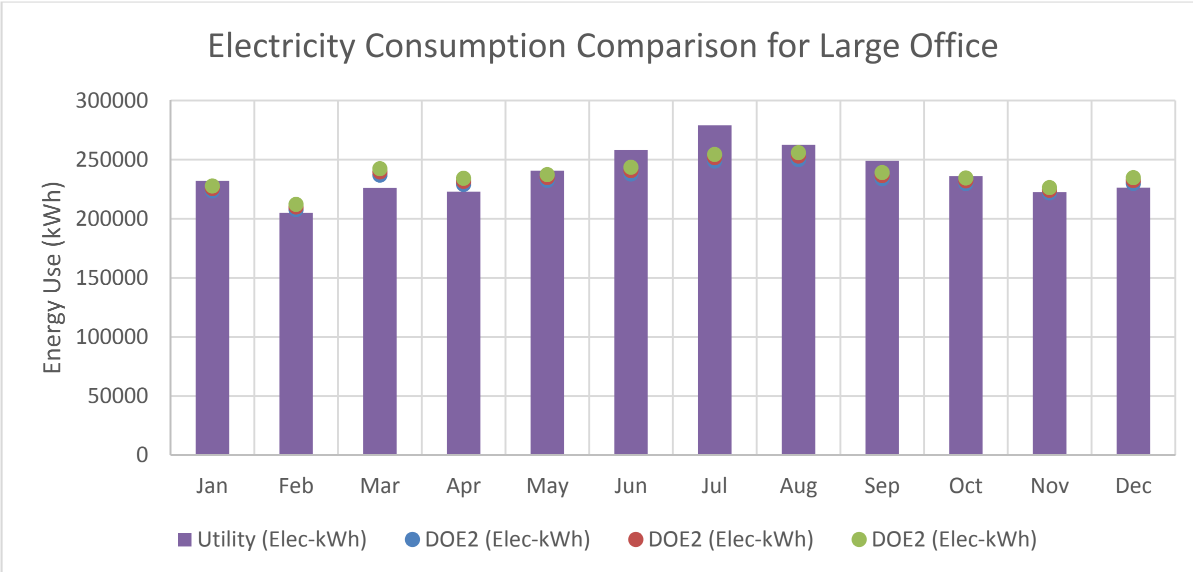


Figure 5-9. Unknown parameter identification result for a large office building



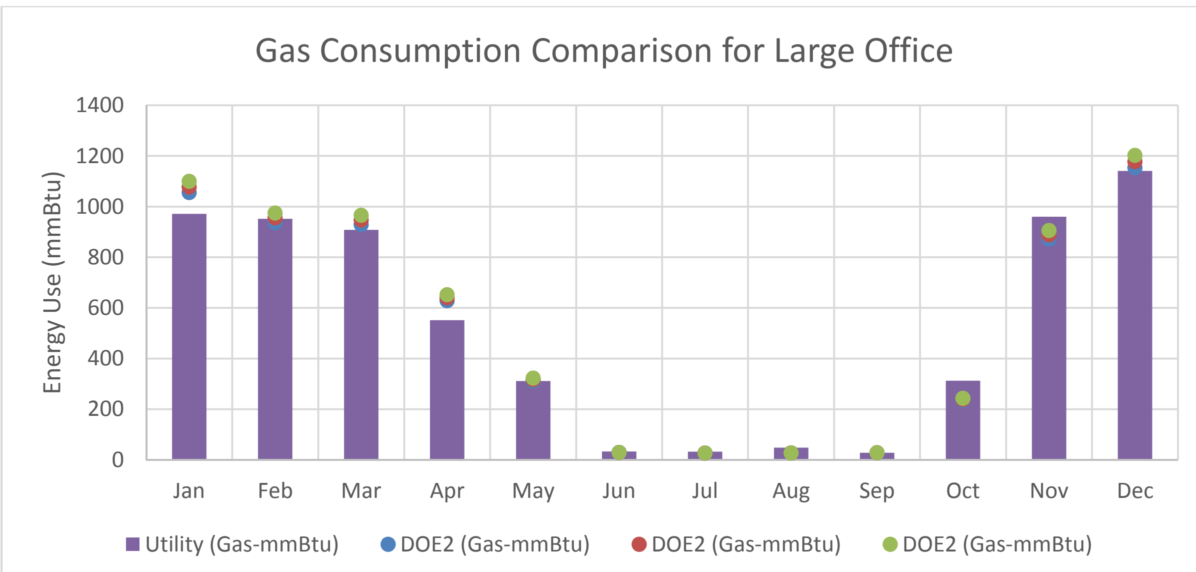


Figure 5-10. After run using predicted values for a medium office building

5.4. Case study of a small residential house

5.4.1. Description of case study

The building assessed in this study was a one-story, family-of-four residence built in 1958 located in a residential neighborhood of West Boulder, Colorado. The four-person home is situated at the end of a cul-de-sac, shaded by large and medium height deciduous trees with additional wind shielding from a perimeter fence bordering most of the property line. The interior consists of 4 bedrooms, 3 bathrooms, and an attached office space typically occupied by 2 to 3 employees. The total conditioned living space is ~3,460 sqft of which ~710 sqft represents the weekday office space. The floor plan in Figure 1 below illustrates the remaining interior spaces and their breakdown by typical use (i.e., office, residential or mixed use). Figure 5-11 shows the exterior, a floor plan of the building, and the energy model screen.

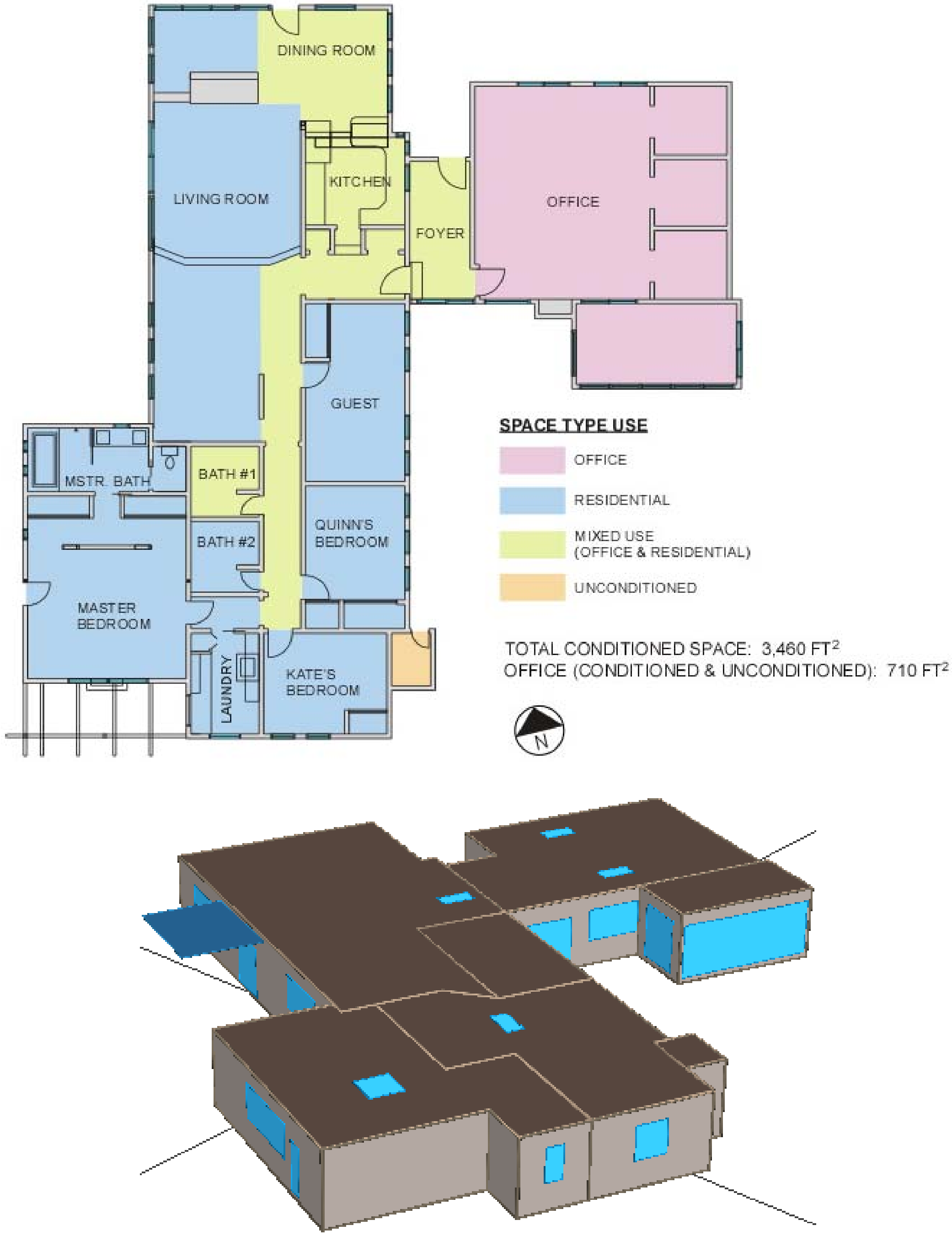


Figure 5-11. Exterior (top) of small residential building and its floor plan (middle) and its eQUEST modeling screen (bottom)

The house is stick frame construction with a bricked lower portion for the exterior walls while the top portion is either a rain screen of horizontal wood plank siding with a vapor membrane or cement board. The exterior facades contain large sections of glazing on the west side to take advantage of mountain views. The total Window to Wall Ratio (WWR) is 28%. The roof is constructed of rolled asphalt atop intersecting gable roofs that contains no attic space beneath. In addition, the roof contains clear and insulated glass skylights in the office, kitchen, hallway and master bedroom spaces. The foyer and office spaces are built on slab-on-grade concrete while the other areas of the house sit above a dirt crawl space with concrete block foundation walls, passively vented to the outside.

5.4.2. Selection of unknown parameters

The following parameters were selected for the identification process. A total of eight parameters were chosen in this case study and are shown in Table 5-3. Among the eight parameters, four parameters that had the most influence on energy consumption were selected. These four parameters are underlined in the table. Values for these parameters were obtained from eQUEST's default values when starting a new project via a wizard. The rest of the values were obtained from a walk-in audit and blueprint. In the minimum and maximum value column, the limits of each parameters which was set by users are shown. These values can be obtained from any reference or based on a user's input. This study considered creating flexibility in putting a limit to each value in consideration of someone with a non-engineering background. After defining the values of these parameters, a sensitivity case varying ±10% was performed. Figure 5-12 shows the sensitivity results for these parameters. The heat input ratio of boiler was the most influential parameter and infiltration was next in energy consumption based on 10% variation. Energy consumption for the window shade coefficient increased as the value decreased.

Table 5-3. Parameters for small residential building (default, minimum & maximum value setting)

Description	Default Value	Min Value	Max Value	Rank	Impact on Energy Consumption difference (%) Per ± 10% variance	Note
Roof U Value (Btu/h-ft2-F)	0.03	0.01	0.5	5	0.35%	
Wall U Value (Btu/h-ft2-F)	0.056	0.01	0.5	3	0.96%	
Window shade coefficient	0.901	0.1	1	8	0.26%	
Infiltration (CFM/ft2)	0.0263	0.01	2.01	2	2.32%	
Light Intensity (kW)	0.3296	0.01	1	7	0.29%	
Equipment Density (W/ft2)	0.642	0.01	1	4	0.47%	
Area/Person (ft2/person)	534.414	100	500	6	0.31%	
Heat Input Ratio of Heating	1.25	1	1.6	1	3.33%	

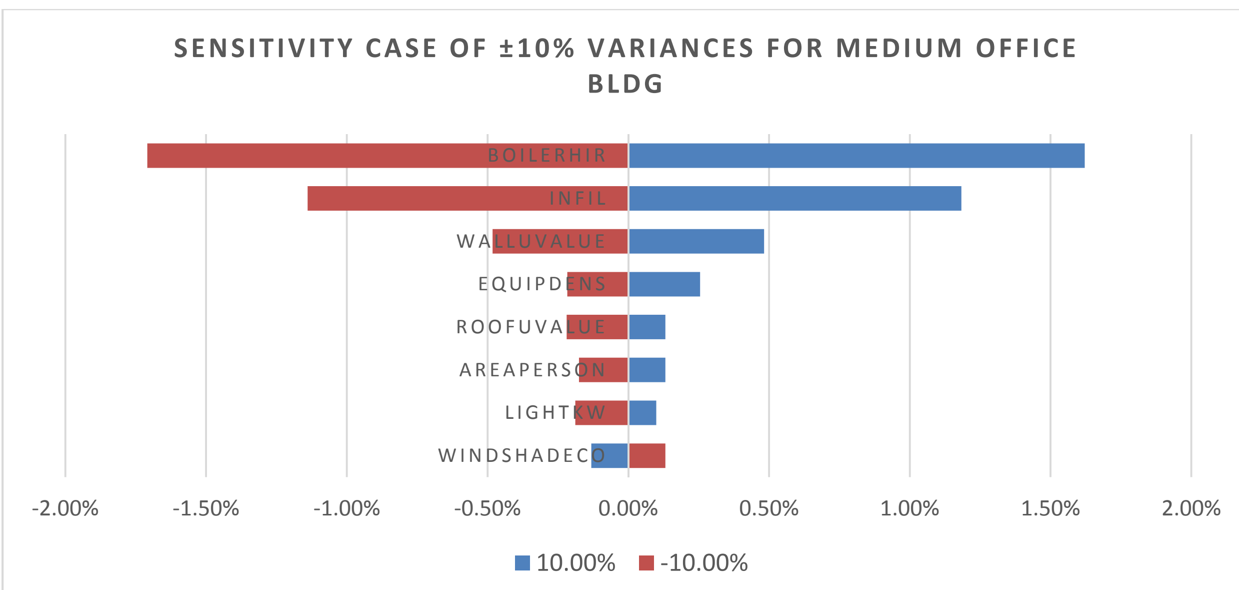
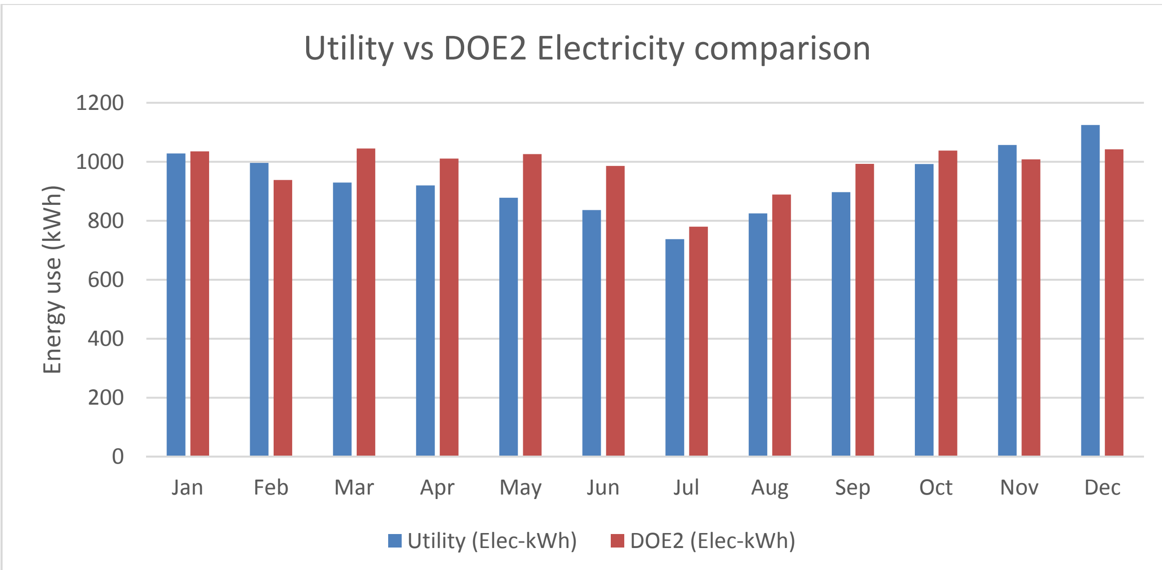


Figure 5-12. Sensitivity result of small residential building after ±10% variations

Figure 5-12 shows the initial energy consumption results for the default values set as in the above table. As shown in the figure, electricity consumption showed a maximum of 20% difference between utility data and initial run, and gas consumption had a maximum of 12% difference. Despite having some slight differences, this model seemed to follow the characteristics of the utility data. Minor changes are needed.



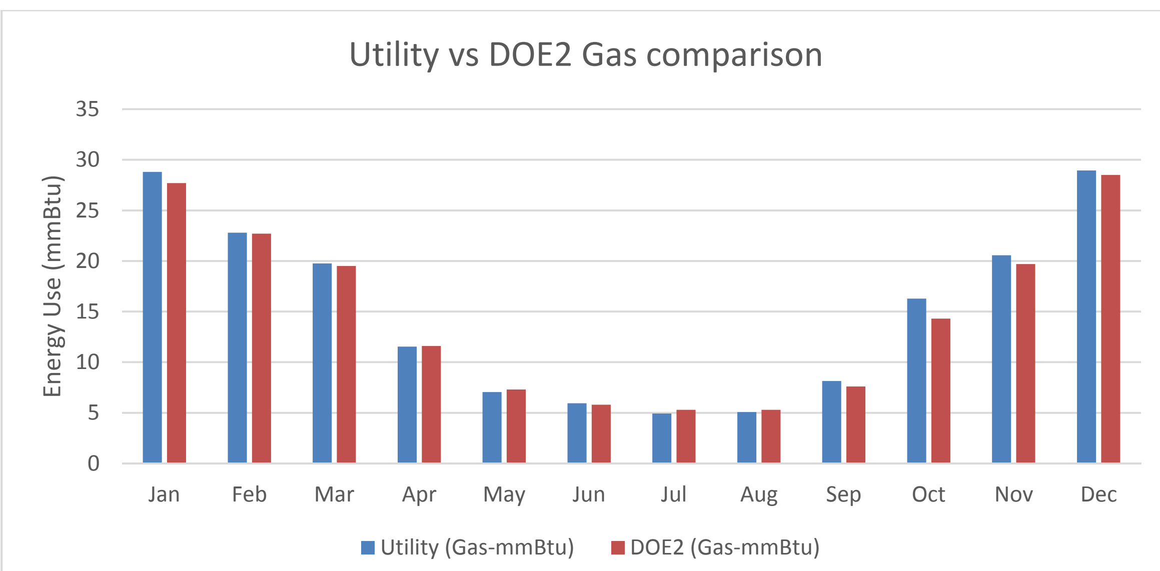


Figure 5-13. Initial run result for large office building

5.4.3. Parameter identification for small residential building

Following the sensitivity case result, this study chose a four unknown parameters identification case to verify the validity. Figure 5-14 shows the posterior distribution of these results. The upper left section shows the u-value of wall, the upper right shows infiltration, the lower left plot shows equipment density results, and the lower right plot shows the posterior of the heat input ratio of the boiler. All four plots suggested the predicted value as the highest occurred bar for each parameter. By using this data from the bar along with its mean, minimum and maximum values, an after-run was performed and the results are shown in Figure 5-15. As shown in the figure, the predicted values of each parameter resulted in closer calibration results. This was especially true for the gas consumption results, and the utility data resides in between the maximum value and mean value result. This suggests that the calibrated value can be found between maximum and mean values. Overall, it took 70 minutes from the start of building the emulator process to the sampling process. This can be considered a fair amount of time for achieving the calibration.

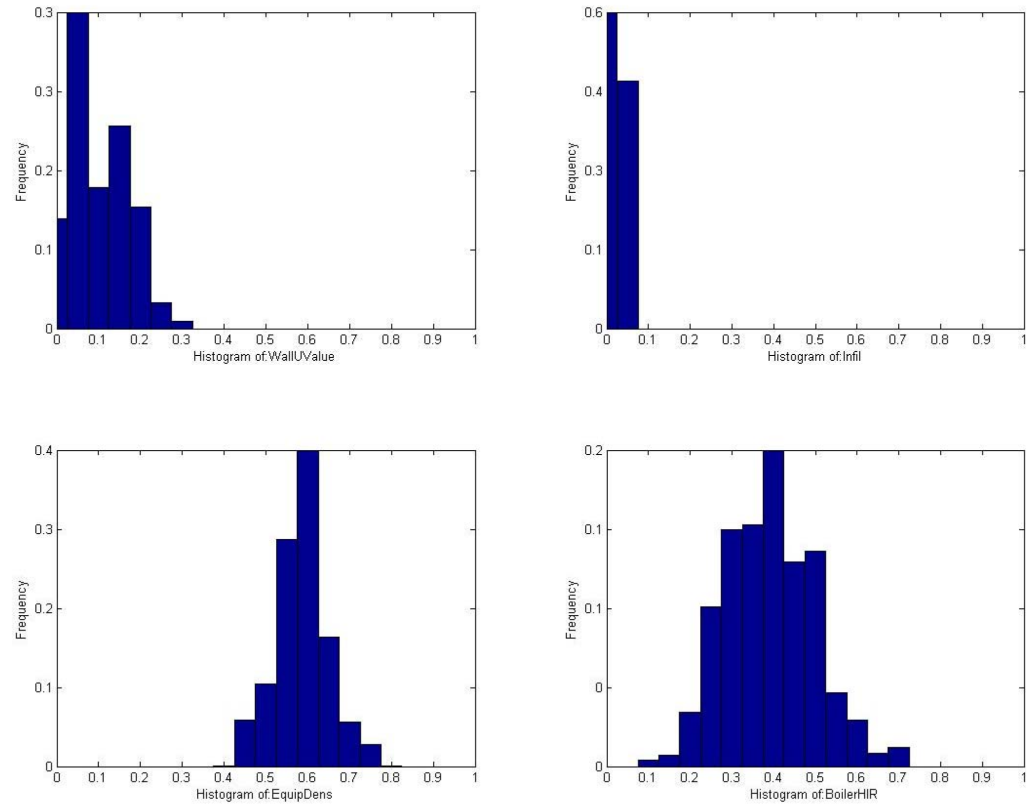
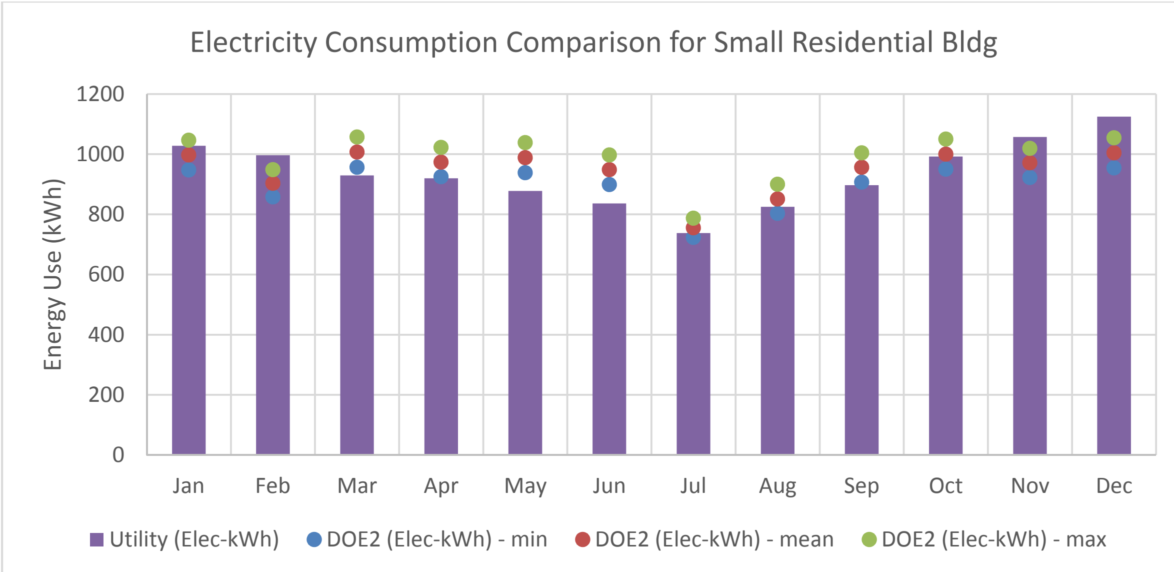


Figure 5-14. Unknown parameter identification result for small residential building



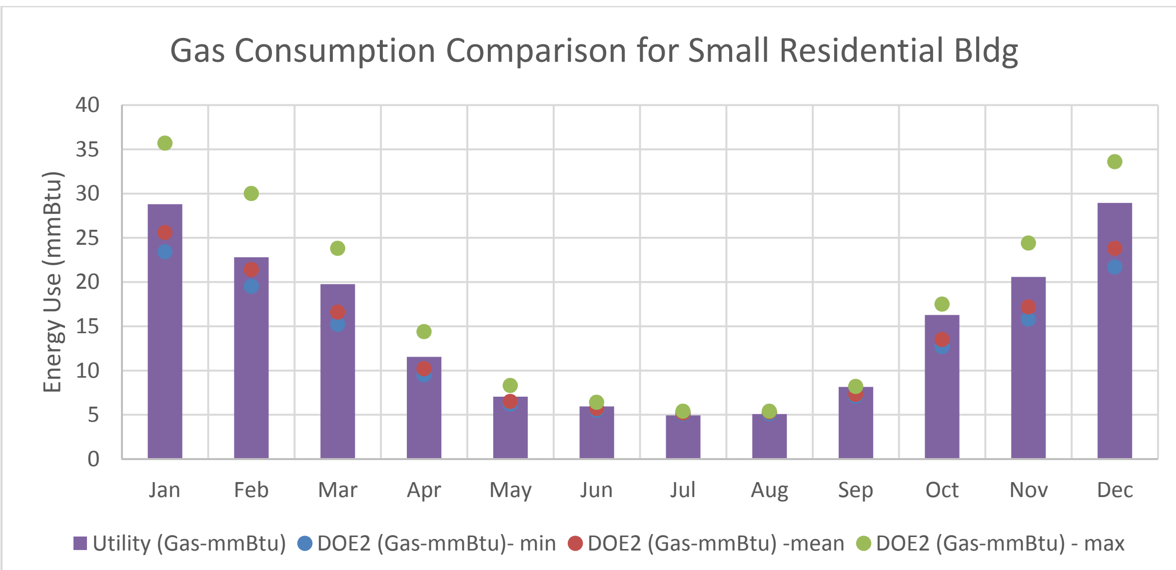


Figure 5-15. After run with using predicted values for small residential building

5.5. Summary and Conclusions

In this chapter, the proposed approach was tested with actual existing building cases to verify validity for unknown parameter identification. Three case studies were selected, consisting of two office buildings (1 medium-17,000 ft2, 1 large-200,000 ft2) and one small residential house.

All of the case studies started with an initial run with values of unknown parameters obtained from eQUEST. After the initial run, our proposed approach was performed to identify each parameter. With the given posterior distribution of each parameter, values with the highest posterior occurrences were chosen and changed the initial run's value and re-ran the simulation to verify if these values were eligible. Each case successfully identified the unknown parameters and therefore achieved calibration after using these obtained values from our approach.

The main concern with the approach was the cpu time that it took to finish the process. In the large office building case, it took over 21 hours to build the emulator. The sampling process in each case study took approximately 90 seconds, so the computational demand for sampling process was low. However, the building emulator process took considerable time when dealing with a larger building. For instance, a single run for a large office building took about a minute to do 1 iteration of simulation. Therefore, obtaining a training set (about 1,000 sets) took a considerable amount of time. Figure 5-16 shows the cpu time and emulator's error comparison result for these case studies. As shown in the figure for the small house and medium office building, there weren't significant differences in CPU time and average error. In fact, all of the cases showed less than 1% average error for the emulators. However, for the maximum error, the large office building case had about 6% error. Considering how much it takes to run the whole process, it is noted that the size of a building influences the practicality of the approach. The size of building means the length of building input parameters. In a large office building, the input file contains about 30,000 lines of inputs. In contrast, the small residential house contained only 3,000 lines. This can influence the overall simulation time since DOE2.2 is a tool that reads all the inputs and calculates according to this information. It is obvious that the larger the input file, the longer it takes to simulate.

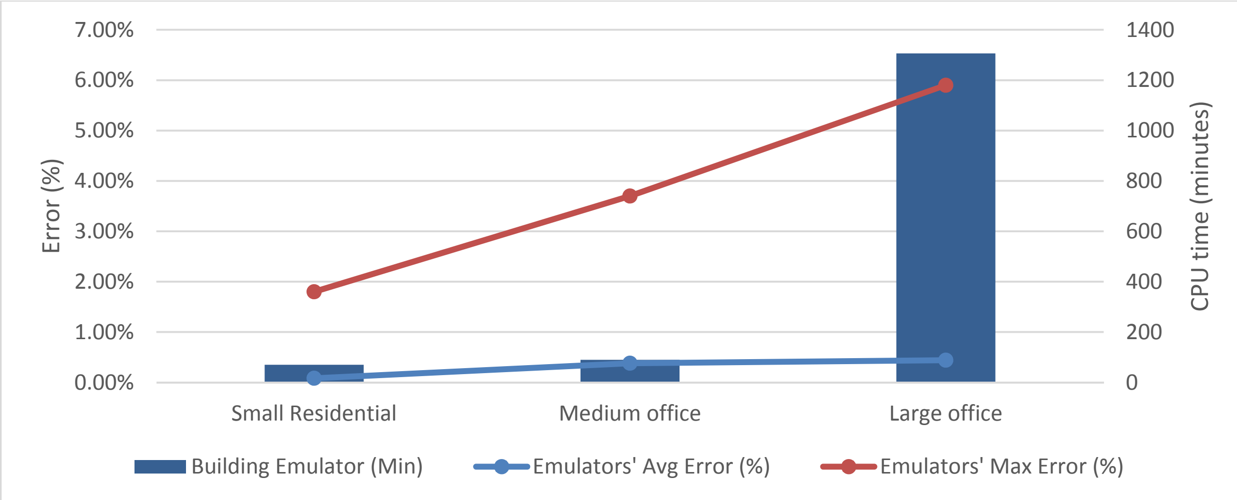


Figure 5-16. CPU time and error comparison for small house, medium and large office bldg model

Until now, this study performed case studies based on existing building models with built eQUEST models. In reality, this cannot be always provide accurate results since sometimes there is a possibility of not going through a thorough energy modeling process. Therefore, a backup plan for this kind of situation needs to be considered. Perhaps building a simplified model and use the proposed approach can be alternative method for parameter identification.

CHAPTER 6: Applications of Bayesian parameter identifications

Until this chapter, this study has focused on developing and validating the proposed Bayesian approach and case studies to verify this approach in reality. The proposed approach has proved to be functioning in several situations for unknown parameter identifications. However, the proposed approach has various applications other than parameter identification problems. One good example will be identifying potential Energy Conservative Measures (ECMs). In this chapter, we will discuss about potential applications that can be applied to the proposed approach. Not only identification of such problems, this study require comparison studies of defining the difference between the Bayesian approach and other regression analysis. Therefore, this study has performed a comparison study of determining the differences between an emulator with linear models and an emulator with Gaussian process. After defining the differences between other regression methodologies, this study has performed simplified model analysis to verify the possibility of the proposed approach to be applied to an energy model with complex geometry by using simpler geometry. At the end, to validate the difference between the traditional Bayesian approach and the proposed approach, this study has performed evaluation study using these two approaches.

6.1. Identification of ECMs (Energy Conservative Measures)

6.1.1. Introduction

After developing a well-calibrated model for existing buildings, energy engineers use this model to evaluate potential energy evaluations. The potential evaluation procedure has a goal of finding potential energy saving, many from what the industry calls Energy Conservative Measures (ECMs) or Energy Efficiency Measures (EEMs). Deriving ECMs is to identify the mixture of various ECMs with the highest energy saving or cost savings and assisting decision makers to define what the most suitable ECMs are for themselves. As anyone can suspect, well-defined energy models have higher accuracy in prediction of energy savings. However, defining the most suitable ECMs also presents many uncertainties due to various characteristics (uncertainties in actual effectiveness of ECMs installment, cost of ECMs, etc.). If the decision factor concentrates on the total cost, or if the decision factor solely concentrates on energy saving, the answer will be very different. Additionally, uncertainties in the actual performance of these ECMs can result in incorrect ECM evaluations.

In current section, this study will concentrate on defining ECMs by using a Bayesian approach. This will include quantifying uncertainties in the ECMs and evaluating the whole process with effective methodology such as using an emulator to reduce CPU burden.

6.1.2. Literature review

To find the potential ECMs can lead to a potential energy savings as well as CO2 emission reduction. Stankovic (Stankovic, et al., 2009) investigated the average energy savings that occurred with refurbished buildings and found out that average of 40% energy have saved and reduction of CO2 emission by 42% average. The Simple Payback Period (SPP) for general building renovation was found to be 7.5 years, school had about 12.8 years and hospitals had 5.3 years because of their 24/7 hour operation.

Kneifel (Kneifel, 2009) studied about estimating life-cycle energy savings, carbon emission reduction and cost effectiveness of energy efficiency measures by using an integrated design approach in new commercial buildings. This study has performed 576 energy simulations with using 16 locations, 12 building types, 3 building designs and built a database to determine life-cycle cost-effectiveness and carbon emissions for different building design. The result of conventional energy efficiency technologies showed that the average energy saving of 20~30% and 40% for some building types and locations. Not only just the energy savings, carbon footprint was reduced 16% on average.

Another study investigated the economic impact of a 5% improvement of household energy efficiency in United Kingdom (Lecca, et al., 2014). The study measured the impact by using simulation models that have increasing degrees of endogeneity with calibrated on a common data set with using a fully specified general equilibrium treatment. The study concluded that the rebound effect is shown to depend on changes in aggregate economic activity, household income and relative prices.

Heo (Heo, et al., 2011) presented a risk analysis method based on calibration of a Bayesian approach for building energy models. By enabling a Bayesian approach and probabilistic results for the energy model, she was able to apply the model to considerations regarding investing in energy conservative measures in existing buildings. The Bayesian approach was then compared to standard practices endorsed by the International Performance Measurement and Verification Protocol (IPMVP, 2010) and ASHRAE guideline 14 (ASHRAE, 2002). The study proposed a calibration and retrofit analysis methodology that can be implemented without expertise and a background in building energy modeling. However, the study also suggested that at the current stage, the proposed method still relies heavily on experts' judgments, especially in the choice of calibration parameters. Additionally, the study was only concentrated on calibration with gas utility data and this can lead to incorrect calibration because the target building also utilized electricity. Furthermore, this study neglected the effect of human behavior related parameters such as schedule of occupancy, lighting, equipment, etc.

6.1.3. Methodology

Normally during the energy audit, numerous numbers of potential ECMs can be found and require same amount of effort to do evaluations on each ECMs. Performing evaluations on ECMs requires time and effort and last but not least, requires general expertise on building energy. In previous chapter, this study has formulated an emulator to alleviate the burden of time and effort. By using similar methodology, this study will perform ECM analysis to find potential ECMs. However, ECMs usually comes with a lot of uncertainties such as quantifying actual improvement of ECMs after installation, determining actual initial cost and determining the effective solutions of ECMs. In order to quantify these, this study have selected Bayesian approach to search for optimized ECMs.

To add, ECM evaluation requires a lot of parameters to account for. In previous chapter, this study have selected partial parameters to formulate an emulator. To determine the possibility of adding additional parameters to an original emulator, this study have performed a brief study of adding a couple of parameters to an original emulator. An original emulator carried 1,000 training sets and we have added two more parameters with 200 training data to determine if this procedure is capable of creating a new emulator. Figure 6-1 shows the RMSE result of addition of 2 variables to an original emulator with 4 unknown variables. To compare the result, 3 cases were created. 3 emulators with 500, 1,000, 2,000 training sets were evaluated the post RMSE after adding 2 variables to an original emulator. As shown in the figure, an emulator with 500 training sets had resulted highest errors among others. And it showed all of emulators seem to result higher RMSE than an original emulator with 4 unknown variables. This is mainly due to the addition of unknown variables. However, despite of increase in error, these emulators still have higher accuracy to be used as an emulator because of lower RMSE. With this result, this study can decide that the addition of parameters to an original emulator is capable.

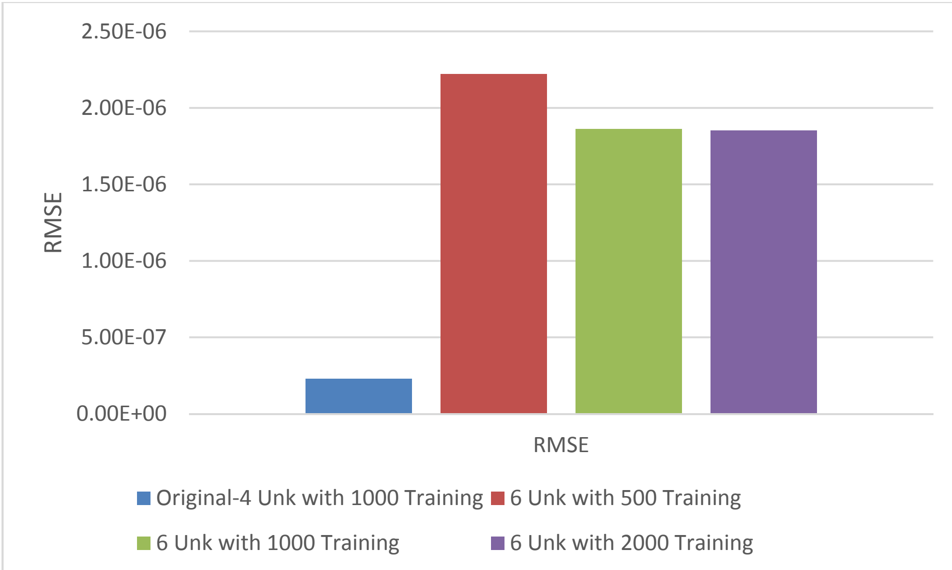


Figure 6-1. Comparison result of RMSE of 2 variables addition to an original emulator for residential building

With above result, this study has selected 6 parameters to be used in the ECM selections. Table 6-1 shows these ECM selections and their property. Before performing any ECM evaluations, this study has quantified uncertainties in these ECMs as shown in Table 6-1. Parameter value column indicates actual parameter's value after the ECM installations. And these values are estimated to carry ±10% uncertainties based on its average value so this study has quantified these values with minimum (-10%) and maximum value (10%) of potential improvements and initial costs that is associated with it. We have given them uniform distribution on these values since actual values can be unknown due to many factors. As well as parameter values, initial cost were estimated with minimum and maximum price with uniform prior distribution.

Table 6-1. Items of ECMs and its properties

			Calibrated	Improved value			Initial Cost ($)	
	Tasks	Parameter Name	Value	Avg	Min	Max	Min	Max
ECM1	Adding insulation	Roof U Value (Btu/h-ft2-F)	0.03	0.0255	0.0230 (-10%)	0.0281 (+10%)	1690.5	1932
ECM2	Adding insulation	Wall U Value (Btu/h-ft2-F)	0.04	0.0340	0.0306 (-10%)	0.0374 (+10%)	1690.5	1932
ECM3	Air-tightening	Infiltration (CFM/ft2)	0.0735	0.0621	0.0559 (-10%)	0.0683 (+10%)	1094.8	1207.5
ECM4	Lighting upgrade	Light Intensity (kW)	0.33	0.2307	0.2076 (-10%)	0.2538 (+10%)	869.4	966
ECM5	Installing smart occupancy adapters	Equipment Density (W/ft2)	0.59	0.5291	0.4762 (-10%)	0.5820 (+10%)	128.8	209.3
ECM6	Boiler upgrade	Heat Input Ratio of Heating	1.25	1.1223	1.2346 (-10%)	1.0101 (+10%)	515.2	579.6

After selecting these parameters, Bayesian approach of finding best scenario for these ECMs was performed. Scenarios that we considered were consist of 4 scenarios. First two consisted of energy savings ($) and CO2 savings ($). Latter two were consisted of Life Cycle Cost (LCC) with or without CO2 credit. For Life cycle cost, Uniform Series Present Worth factor (USPW) which utilizes life cycle year (n) and discount rate (d) was used (10 years of life cycle and 5% of discount rate was assumed). And these scenarios are described in below. And data for utility price, CO2 price, CO2 credit price is shown in Table 6-2.

$$Scenario\ 1 = \sum_{i=1}^{12} Energy\ Saving * Utility\ Price_{gas,elec}(\$)$$

$$Scenario\ 2 = \sum_{i=1}^{12} Energy\ Saving * CO_2\ production_{gas,elec}(\$)$$

$$Scenario\ 3 \ = \ Initial\ Cost(\$) + EnergyCost\left(\frac{\$}{year}\right) * USPW$$

$$USPW = \frac{1-(1+d)^{-N}}{d}$$

$$Scenario\ 4\ =\ Initial\ Cost(\$) + \left[EnergyCost\left(\frac{\$}{year}\right) - CO_2Credit\ \left(\frac{\$}{year}\right)\right] * USPW$$

Table 6-2. Utility price, CO2 price, CO2 credit prices for US

	Price per unit	Reference
Electricity price per kWh	\$0.12	(U.S. DOE, 2014a)
Gas price per therm	\$0.72	(MGE, 2013)
CO2 (lbs) reduction price per kWh from Electricity	\$1.50	(EPA, 2014)
CO2 (lbs) reduction price per MBtu from Gas	\$117.08	(U.S. DOE, 2014b)
CO2 credit price per tons	\$30	(Nordhaus, 2008)

6.1.4. Results

First, this study has tested the most cost saving ECMs and ranked them by their amount of savings and Table 6-3 shows this result. After 1,000 sampling runs from Bayesian approach, this study have determined that equipment density reduction had the highest cost savings based on its low installment cost. On the contrary to equipment density, adding insulation to wall had the least impact on cost savings. This is probably due to higher cost that was involved in installment of these features. **Figure** 6-2. Posterior of LCC **prediction** with CO2 **credit of 6 ECMs for residential building (continued)**

shows the posterior of all of scenarios result of these six ECMs. X axis represents the life cycle cost after 10 years. As shown in the figure, despite of uncertainties involved in these ECMs, Bayesian approach will derive to certain optimized life cycle costs.

Table 6-3. Ranking of 6 ECMs by each scenarios

	Scenario 1 (Energy Savings)			Scenario 2 (CO2 Savings)			Scenario 3 (LCC Savings)			Scenario 4 (LCC with CO2 credit)		
	Rank	Savings ($)	Savings (%)	Rank	Savings ($)	Savings (%)	Rank	Savings ($)	Savings (%)	Rank	Savings ($)	Savings (%)
ECM1	5	304.2	11.1%	2	4227.6	10.8%	6	1551.0	7.3%	4	8476.2	40.1%
ECM2	3	304.5	11.1%	5	4187.1	10.7%	5	1559.1	7.4%	5	8307.6	39.3%
ECM3	2	305.0	11.1%	6	4180.7	10.7%	3	2129.9	10.1%	3	8822.1	41.7%
ECM4	1	349.7	12.8%	1	4804.6	12.3%	4	2079.8	9.8%	6	8192.7	38.8%
ECM5	6	295.2	10.8%	4	4221.3	10.8%	1	3089.6	14.6%	1	10012.4	47.4%
ECM6	4	304.3	11.1%	3	4227.6	10.8%	2	2723.3	12.9%	2	9650.1	45.7%

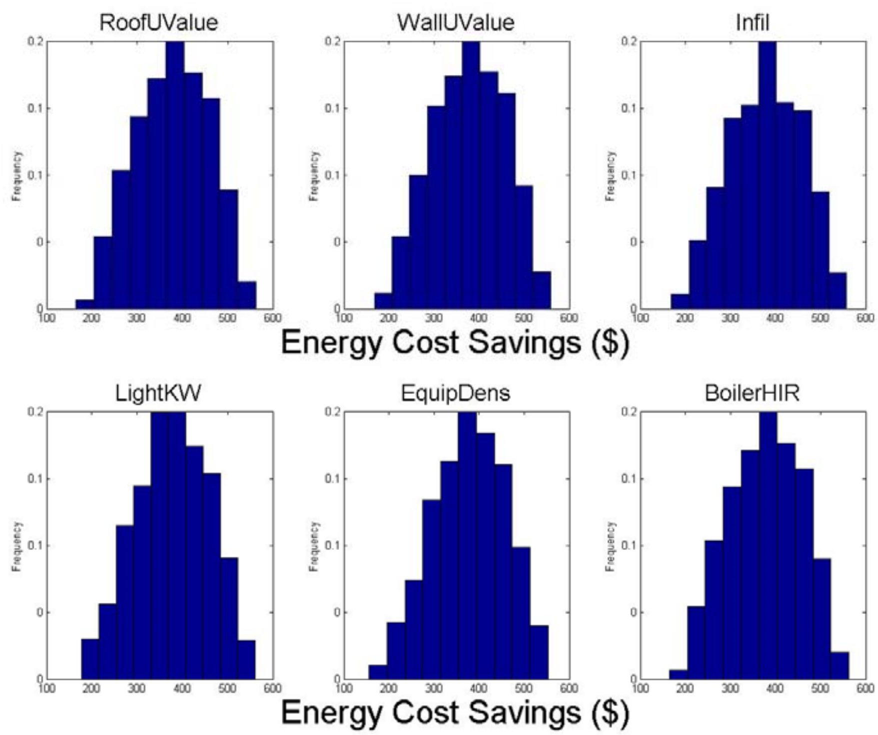


Figure 6-2. Posterior of LCC prediction with CO2 credit of 6 ECMs for residential building

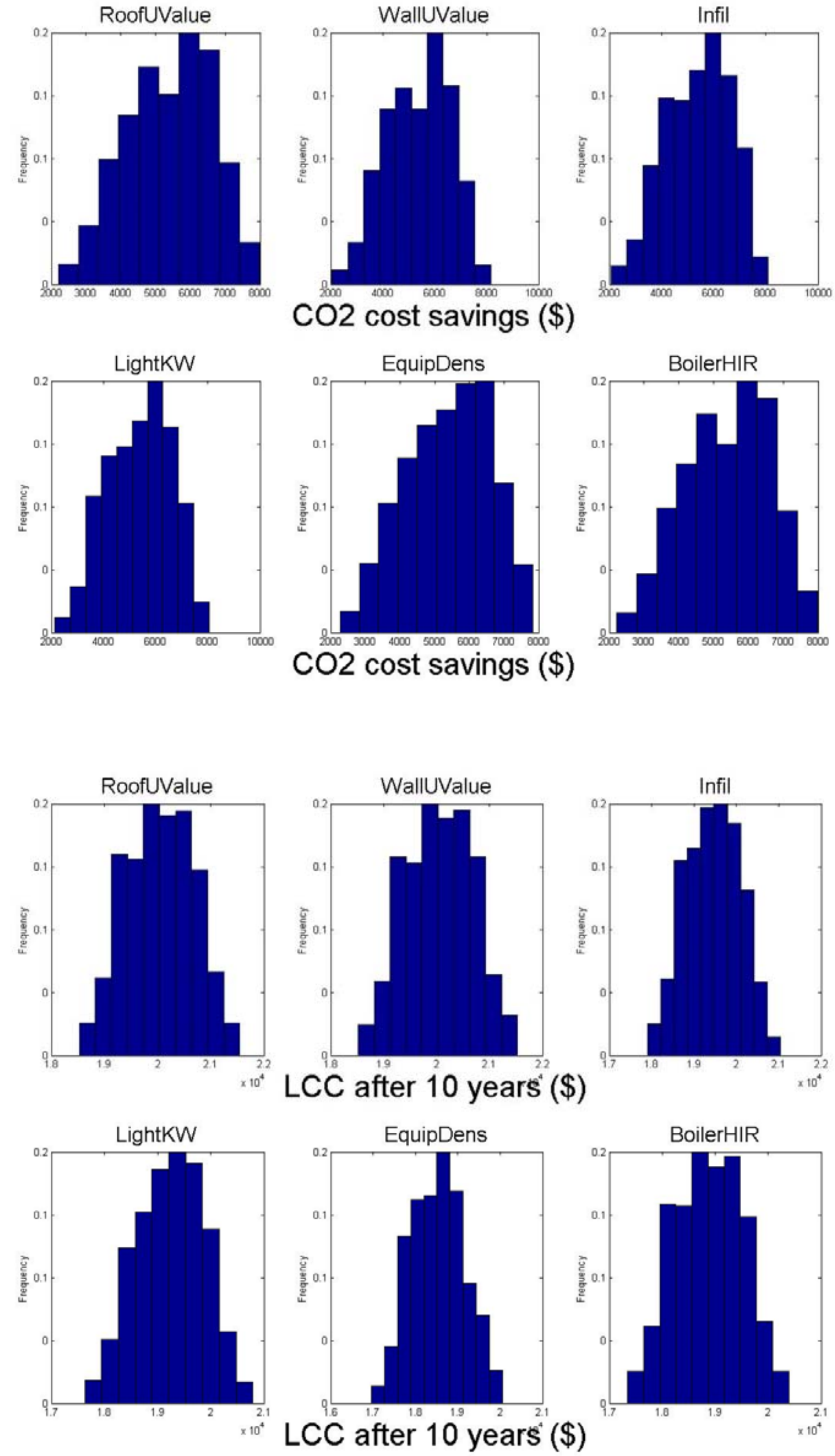


Figure 6-2. Posterior of LCC prediction with CO2 credit of 6 ECMs for residential building (continued)

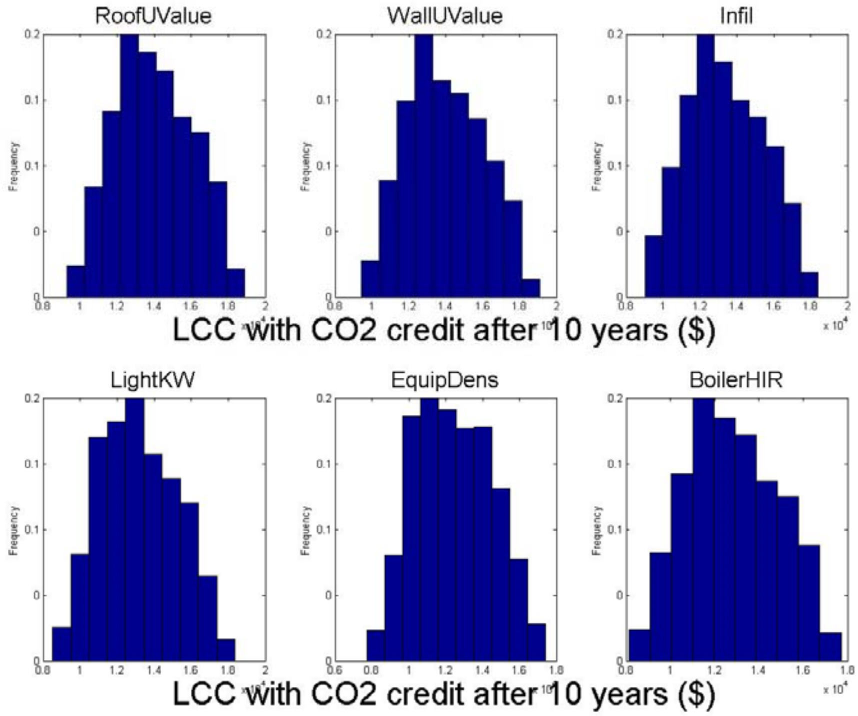


Figure 6-2. Posterior of LCC prediction with CO2 credit of 6 ECMs for residential building (continued)

At certain scenarios, an owner of the building might have limited budget to work on these ECMs. Due to this limitation, there is a need for a decision making procedure for these owners to choose among various ECMs. This study has created a subroutine that will derive optimized result with given budget amount. In this scenario, this study has selected three budget options to select these ECMs. Table 6-4 represents Optimized ECMs selection based on each budget limitations. For given low budget, combining ECM5 and ECM6 as the selections for ECMs. As shown in the table, the successful candidate of ECMs were combining ECM2, ECM3, ECM4, ECM5, ECM6 for scenario 4. Estimated savings were 33.3% from overall life cycle cost for 10 years.

Table 6-4. Optimized ECMs selections based on budgets for residential buildings

	Budget < $1,500			Budget < $3,000			Budget < $5,000		
	Recommended ECMs	Savings ($)	Saving (%)	Recommended ECMs	Savings ($)	Saving (%)	Recommended ECMs	Savings ($)	Saving (%)
Scenario 1	ECM(4+6)	$237	8.7%	ECM(3+4+5+6)	$344	12.6%	ECM(2+3+4+5+6)	$360	13.1%
Scenario 2	ECM(4+6)	$3,402	8.7%	ECM(3+4+5+6)	$4,798	12.3%	ECM(2+3+4+5+6)	$5,044	12.9%
Scenario 3	ECM(5+6)	$516	2.4%	ECM(3+4+5+6)	$1,178	5.6%	ECM(2+3+4+5+6)	$1,205	5.7%
Scenario 4	ECM(4+6)	$4,202	19.9%	ECM(3+4+5+6)	$6,735	31.9%	ECM(2+3+4+5+6)	$7,048	33.3%

	Budget < $1,500			Budget < $3,000			Budget < $5,000		
	Recommended ECMs	Savings ($)	Saving (%)	Recommended ECMs	Savings ($)	Saving (%)	Recommended ECMs	Savings ($)	Saving (%)
Scenario 1	ECM(4+5)	$172	6.3%	ECM(4+5+6)	$313	11.4%	ECM(3+4+5+6)	$344	12.6%
Scenario 2	ECM(6)	$2,277	5.8%	ECM(4+5+6)	$4,290	11.0%	ECM(3+4+5+6)	$4,798	12.3%
Scenario 3	ECM(6)	$23,676	-12.0%	ECM(5+6)	$22,919	-8.5%	ECM(3+4+5+6)	$22,918	-8.4%
Scenario 4	ECM(6)	$21,039	0.4%	ECM(4+5+6)	$17,980	14.9%	ECM(3+4+5+6)	$17,361	17.9%

6.2. Comparison to other regression approach

6.2.1. Introduction

Regression analysis can be a statistical tool for the investigation of correlation between input variables and output variables (Sykes). This means when using a regression analysis, the investigator seeks to find the causal effect of one variable upon another and associated consequences. In our case, we are trying to estimate the energy consumption with using physical and behavioral parameters of buildings. There can be various approaches to do these. One simple example is using the linear regression model. The linear model assumes very basic function of following equation where a is the linear term that correlates x and y, and b is the intercept term.

$$y = ax + b$$

With using this simple form, one can expand it to create a formula of correlating building inputs and energy consumption. In this study, we will try to utilize this linear regression model to replace the Gaussian process model.

6.2.2. Literature review

O'neil (O'neil, et al., 1991) studied about relationships between the impact of six building parameters and heating & cooling of the building. Sensitivity analysis was carried out to obtain the correlations among building parameters and identified parameters with most influence on building energy loads.

In Signor's study (Signor, et al., 2001), 10 building parameters were chosen to calculate the energy consumption of buildings with 14 Brazilian weather files. This study has created one equation for each weather sites to obtain consumption data. In the result, the coefficient of determination (R^2) was varying from 0.986 to 0.996. However, this study only included building geometry parameters and not covering input variables from HVAC.

Westphal (Westphal, et al., 2007) utilized a regression analysis to formulate equations to estimate the electric energy consumption as a function of building parameters. These equations can be presented as simplified tools to support the designer as well the simulationist. A generated simplified equation based calculation had coefficient of determination equal to 0.9730 while covering 17 parameters.

6.2.3. Methodology

To investigate the difference between establishing an emulator with linear model or with a Gaussian process, this study will utilize the four different linear models to develop an emulator.

The number of training sets that will be used for these linear models are 1,000 sets and it was tested with different 1,000 testing sets. These sets were all created with samples from Latin Hypercube sampling method. The linear model, there were 4 different methods. For 'linear' model, it contains an intercept and linear terms for each predictor. For 'interactions', it contains an intercept, linear terms and all products of pairs of distinct predictors (no squared terms). In 'purequadratic' model, it contains an intercept, linearterms and squared terms. 'quadratic' term contains an intercept, linear terms, interactions, and squared terms. Assuming it is 2 dimensional problem, the equation for each term is described below.

$$
\begin{aligned}
&\text{'linear':} && y = \alpha x_1 + \beta x_2 + \omega \\
&\text{'interactions':} && y = \alpha(x_1 * x_2) + \beta x_1 + \gamma x_2 + \omega \\
&\text{'purequadratic':} && y = \alpha x_1 + \beta x_2 + \gamma x_1^2 + \delta x_2^2 + \omega \\
&\text{'quadratic':} && y = \alpha x_1 * \beta x_2 + \gamma x_1^2 + \delta x_2^2 + \varepsilon(X_1 : X_2) + \omega
\end{aligned}
\tag{1.1}
$$

Where, ω is the intercept for each model and $\alpha, \beta, \gamma, \delta, \varepsilon$ are all constant values for each terms.

By using these linear models, this study will test the performance when used in our approach.

6.2.4. Results

Figure 6-3 represents the root mean square error that was encountered during 1,000 testing sets and its computing time while building an emulator. As shown in the figure, Gaussian process had the most CPU time but least percentage error. This can be interpreted as the Gaussian process model has the higher computing burden but carries highest accuracy among these five models.

Among four linear models, as suspected, ‘linear’ model had the highest errors and ‘quadratic’ was the most accurate model. Figure 6-4 shows the resulted scatter plots of January gas consumption testing points with prediction points of these emulators. As shown in the result, the ‘linear’ model showed the most points with failed prediction and the Gaussian Process model had the highest accuracy in predicting these testing points. ‘Quadratic’ had the most accuracy among these linear models but it had failed predict some of testing points.

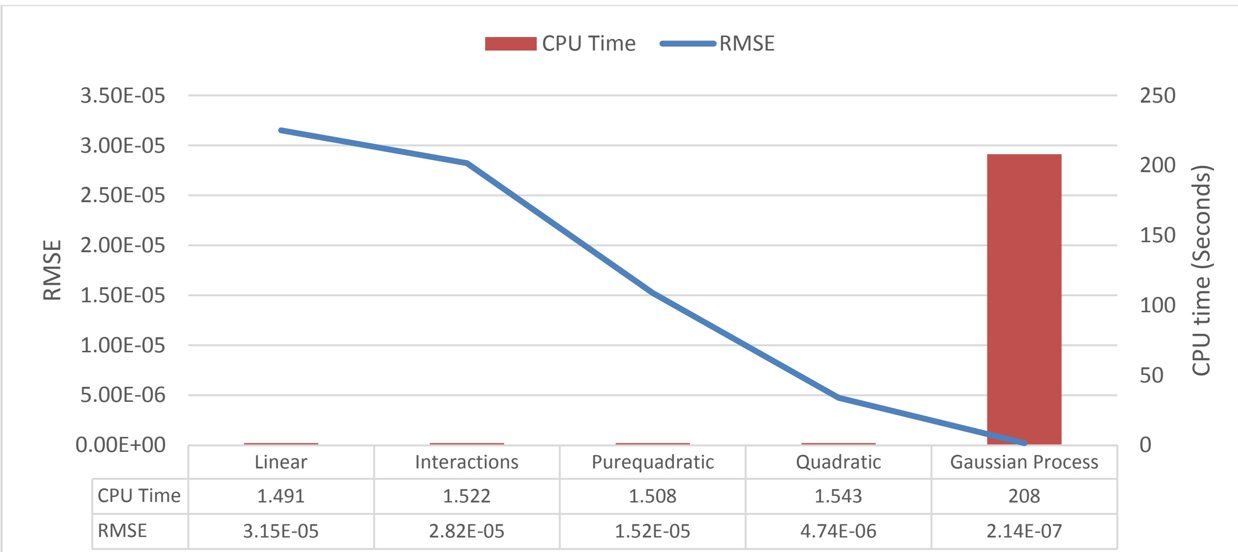


Figure 6-3. Percentage error for linear model and Gaussian process emulator

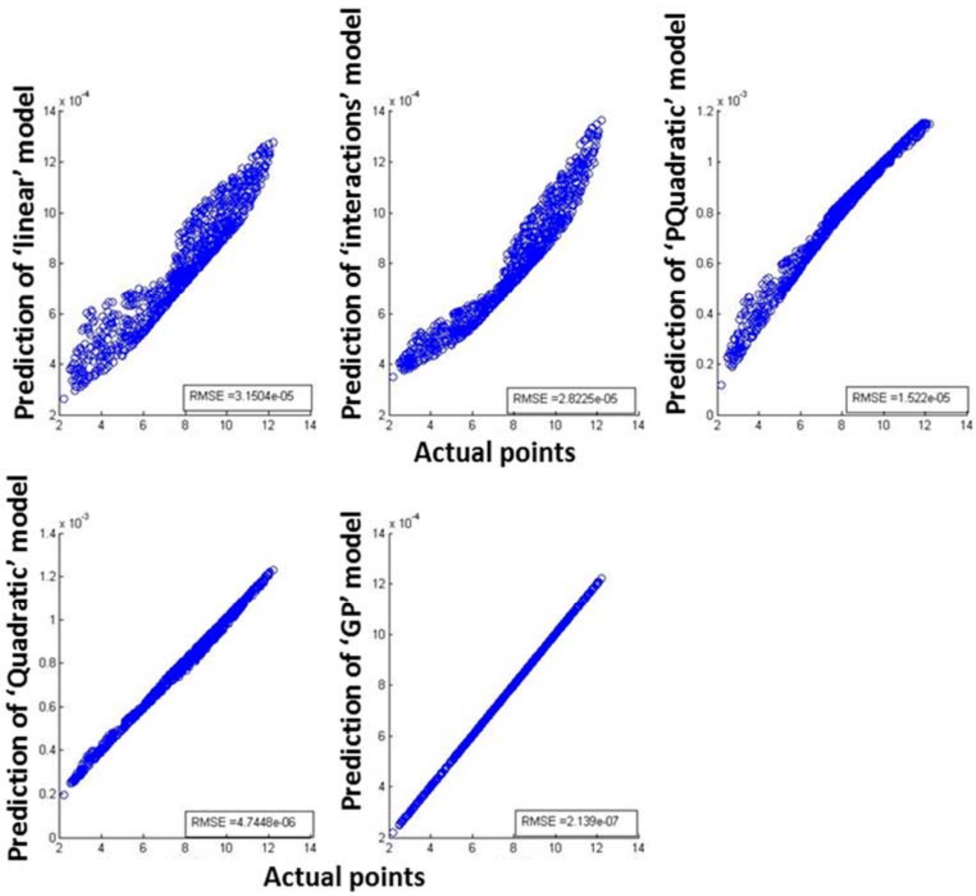


Figure 6-4. Scatter plots for January gas consumption testing points Vs predictions in different emulators

To verify these model's predictions as an emulator, this study has performed Bayesian approach to sample posterior distribution. Then to determine the accuracy with these posterior results, PRMSE was calculated. Figure 6-5 shows the result of PRMSE and CPU time of these models. As shown in the figure, linear models have limitations on deriving posterior distribution with certain accuracy. This kind of result derived from the higher maximum error compared to Gaussian process. In the figure, CPU time of sampling procedure is shown as well. Gaussian Process had about 38 seconds to perform 1,000 samplings and linear models took about 10~13 seconds to sample. Despite of less amount of CPU time needed for the sampling procedure, it shows Gaussian process had the most accurate result. Figure 6-6 shows the posterior distribution result for linear models and Gaussian process model. Dotted line within the plots indicate the target value that the approach is trying to identify. As shown in the result, 'Pure Quadratic' and 'Gaussian Process model was able to show highest posterior occurrences on the target values. However, as indicated in Figure 6-5, PRMSE was higher than that of Gaussian Process.

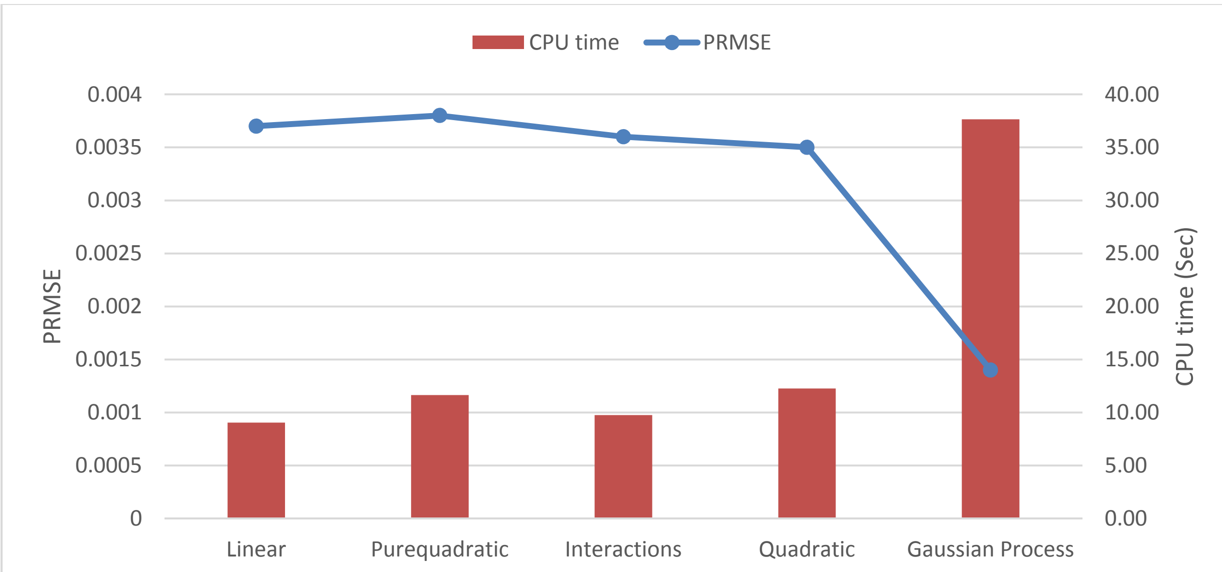


Figure 6-5. Posterior Root Mean Square Error and CPU time for sampling per 1,000 samples

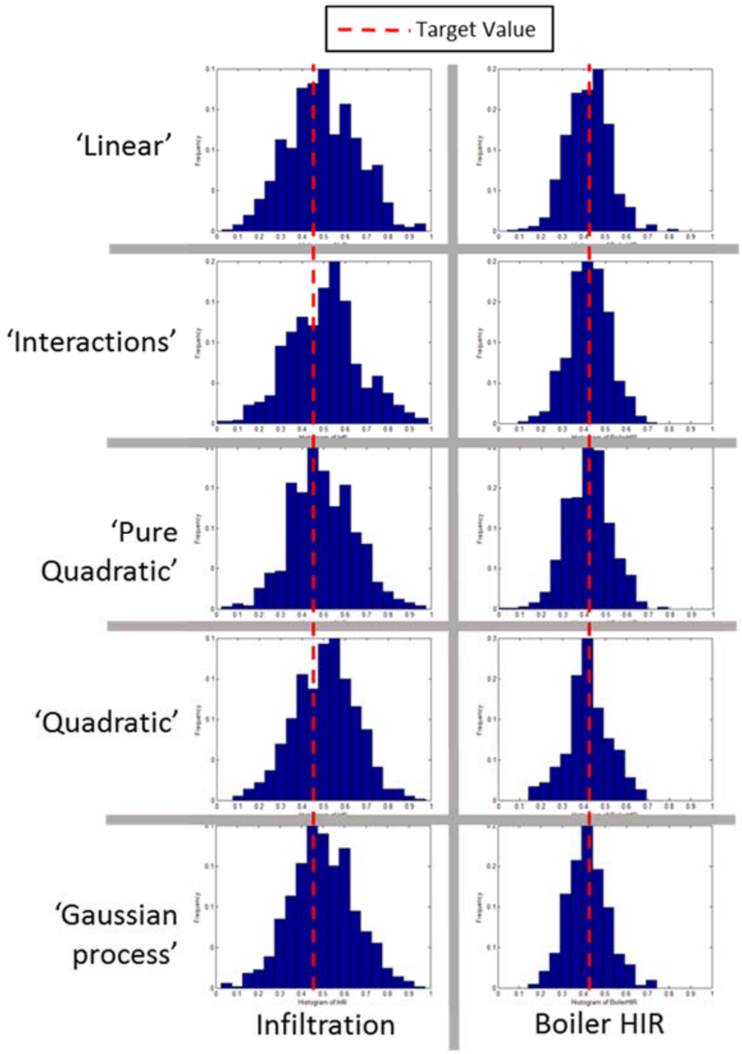


Figure 6-6. Posterior distribution from linear models and Gaussian Process model

6.3. Simplified model analysis

6.3.1. Introduction

Until now, there have been a lot of approaches to accomplish simple and succinct energy simulations. During a building energy modeling process, much effort, time and expertise is required. This is because energy simulation takes more than 5,000 parameters to do a single run. This is why a lot of efforts are required to reduce the overall modeling time as well as simulation time. For instance, Pavlak (Pavlak, 2014) used ISO 13790 to model the building instead of actual energy simulation model to reduce computing burden. To verify the validity of using simplified geometry instead of complex geometry in the energy modeling, this study will utilize energy model with simplified geometry to be used in our approach.

6.3.2. Literature review

Lavigne (Lavigne, 2009) developed and implemented a calibration method to DOE2.1E based building energy simulation software which assists the user with built-in engineering rules as well as optimization algorithm based on Marquardt-Levenberg non linear least square method. This study has performed 2 case studies to represent the algorithm's functionalities. At the end, maximum error on monthly electricity bill was reduced from 143% to 14% for the first case and 40% to 11% for second case.

Urban (Urban, et al., 2007) proposed simplified user interface as well as quick energy model to be used in early stage simulation. He achieved simplified modeling by minimizing the modeling process, specifically the user-interface. The sole purpose is to gain access to non-technical users. The tool that was developed within this study was called 'MIT Design Advisor' and author suggests that it is simplified tool for quickly exploring the energy requirements of competing design concepts.

Valade (Valade, 2009) developed a simplified and concise building simulation model for a preliminary design or embedding into adaptive control systems. The study developed a simplified single-zone model and compared with the real building's performance. The annual loads of the building at a difference of 0.14%, followed by the eQUEST model 4.2% and original DOE2.2 model with 13.3% difference.

6.3.3. Methodology

For simplified model analysis, this study has selected 4 shapes to compare. All buildings were built on 25,000 sqft floor area and buildings' specification is shown in Table 6-5. All of the physical and behavioral properties are the same except for its shape. Especially, the only difference between 'Square' and 'Rectangular' models are the aspect ratio. Figure 6-7 shows the exterior appearances for these buildings and Figure 6-8 shows its energy consumption result. In the energy consumption figure, electricity result is shown in above and gas consumption result is shown in below. As shown in the figure, 'L-Shape' building had the most gas energy consumption. Due to its complex geometry, wall area and window area are bigger than that of other buildings. This means that larger exterior surfaces will absorb or emit the heats from in and out. This will cause the heating and cooling load to increase from the rest of building shape. As shown in the table below, only difference in energy consumption were all related to heating and cooling system.

Table 6-5. Building properties of 4 different simplified model buildings

Parameter name	L-Shape	Square	Rectangular	Rectangular
Aspect Ratio	NA	1	1.5	2
Total Floor Area (ft2)	25,000	25,000	25,000	25,000
Floor to Ceiling Height	9	9	9	9
Building shape	L-shape	Square	Rectangle	Rectangle
Aspect Ratio	1	1	1	1
Light density (W/Sqft)	1.49	1.49	1.49	1.49
Equip density (W/sqft)	0.75	0.75	0.75	0.75
People	62.5	62.5	62.5	62.5
WWR	Flr-flr (40%), Flr-Ceil(53.3%)	Flr-flr (40%), Flr-Ceil(53.3%)	Flr-flr (40%), Flr-Ceil(53.3%)	Flr-flr (40%), Flr-Ceil(53.3%)
Window Area (ft2)	5121.42	4263.17	4352.45	4523.34
Wall Area (ft2)	7756.98	6469.63	6603.55	6859.86
Space Heating (MBtu)	197.7	163.2	159.6	160.8
Space Cooling (MBtu)	110.5	106.2	105.1	104.9
Vent Fans	62.8	58.7	58.6	59.4
Location	Boulder, CO	-	-	-

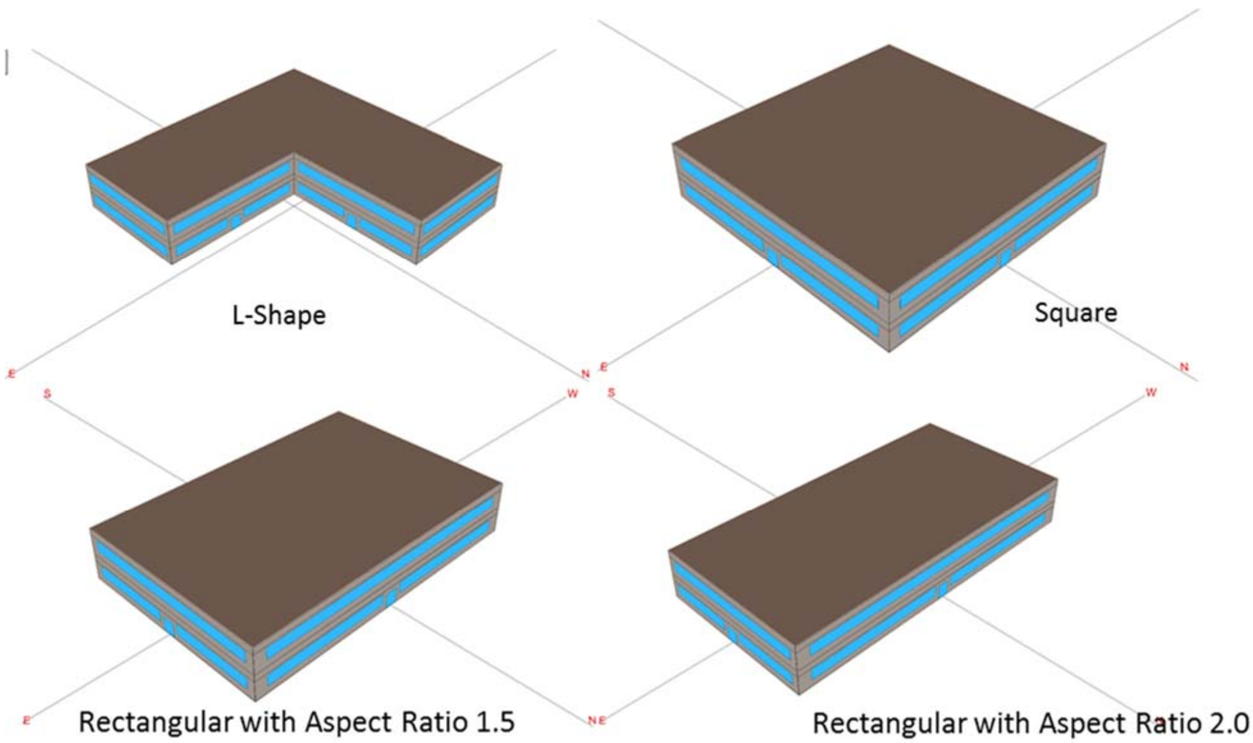
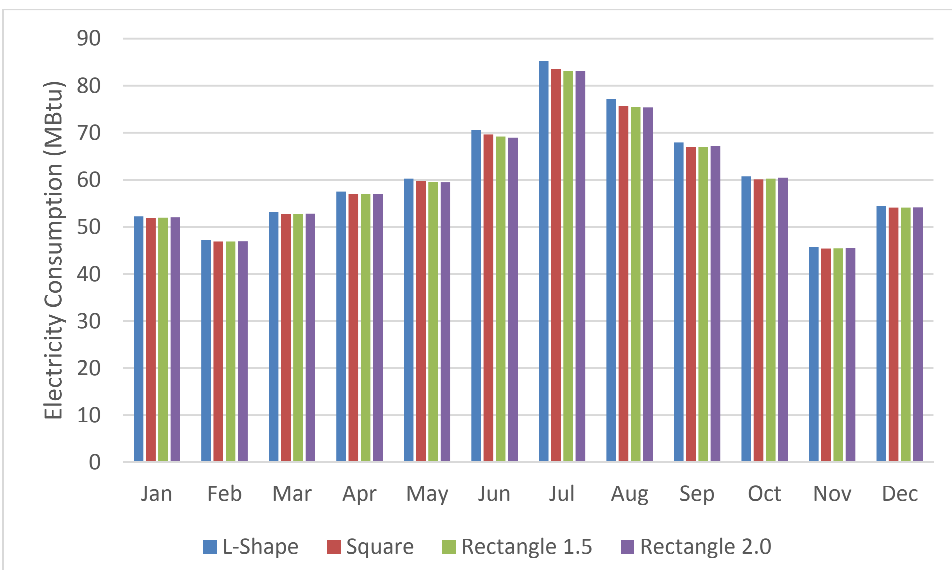


Figure 6-7. Exterior appearances of 4 different simple building models



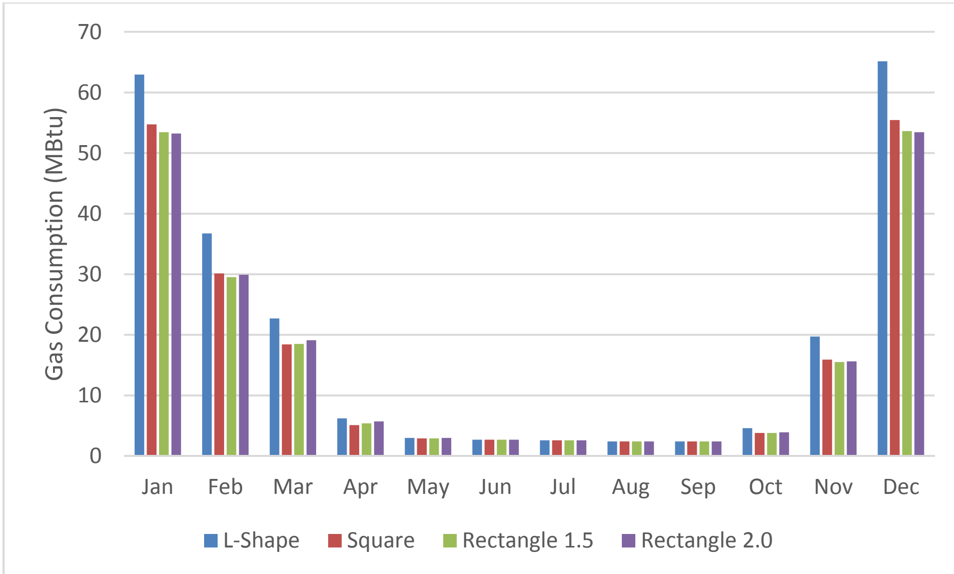


Figure 6-8. Monthly energy consumption for 4 simple building models

In this section, this study is trying to use different models to identify unknown parameters from the original model. Specifically, this study is trying to identify 2 parameters (infiltration, heat input ratio of boiler) of an 'L-shape' building, which has complex geometry, using other simplified models, 'rectangular' or 'square'. This study will create Bayesian approach with using 'Square', 'rectangular' shaped models and formulate an emulator to sample the distribution at the end.

6.3.4. Results

Figure 6-9 represents the result of PRMSE for these four models. As shown in the figure, PRMSE tends to decrease as the overall wall and window area gets closer to 'L-shape' building. There may be a possibility that this decrease in error derived from area of exterior components. This can be noticed in the figure, as the PRMSE increases toward the 'L-Shape' building the accuracy tends to increase as well.

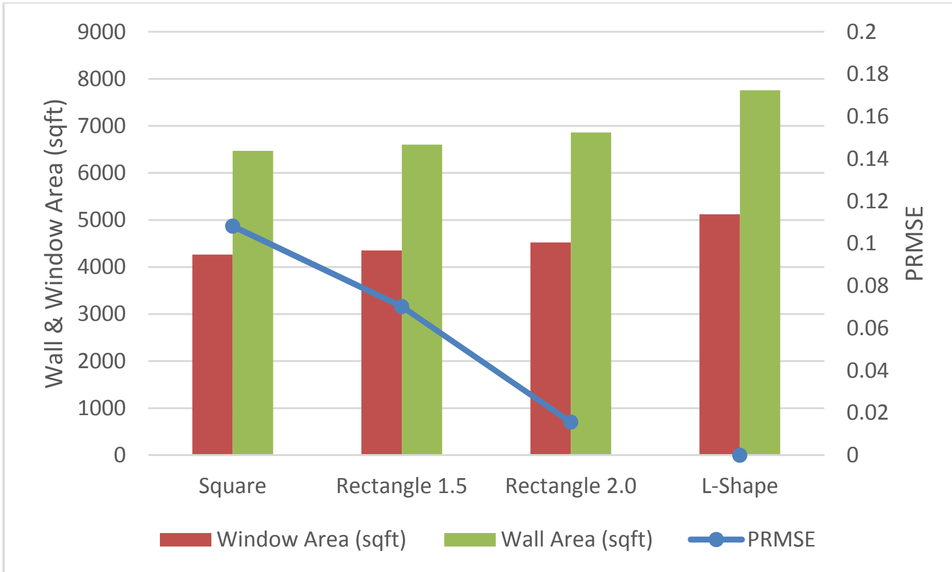
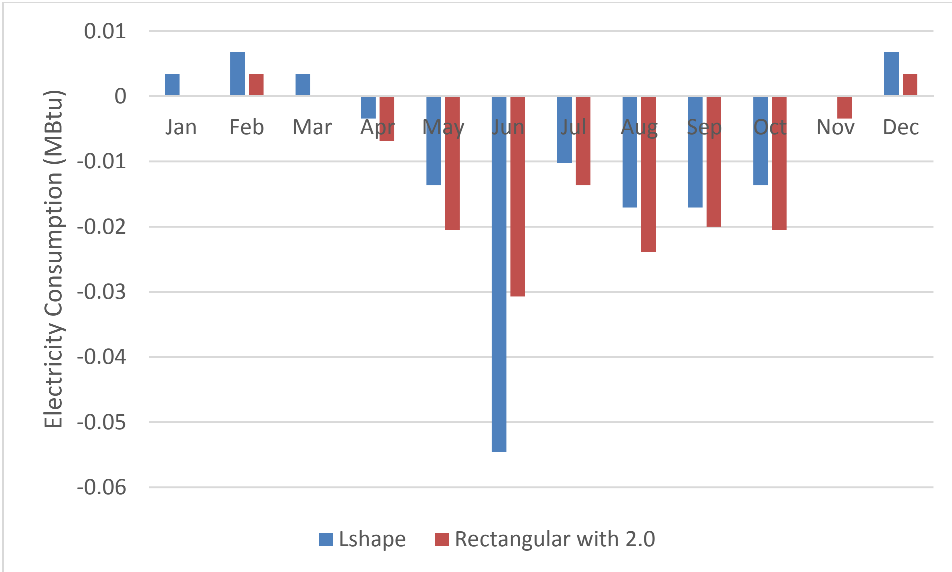


Figure 6-9. Posterior Root Mean Squared Error result of simplified models

To verify the impact of variances in value of a parameter, this study has performed sensitivity analysis on 'L-Shape' building and 'Rectangular with aspect ratio of 2.0'. These two shapes were selected to vary 0.0014 (approximately 10% of original value) to see the difference in energy consumptions. Figure 6-10 represents the electricity and gas difference after 10 % variation in infiltration rate for both of the buildings. As shown in the figure, energy consumption gradually increased in heating season and gradually decreased in cooling season after 10% variation to the infiltration rate. And the differences derived from energy consumption between these two models didn't showed a large changes after 10% increase in the value. This indicates that the despite of their exterior appearance difference, the energy consumption will be similar if the exterior area is similar to each other. Therefore, the possibility of applying simplified modeling strategy to the proposed Bayesian approach is feasible. However, the simplified approach still needs a further study on the approach of simplifying the building exterior as well as case studies that will support and validate the theory.



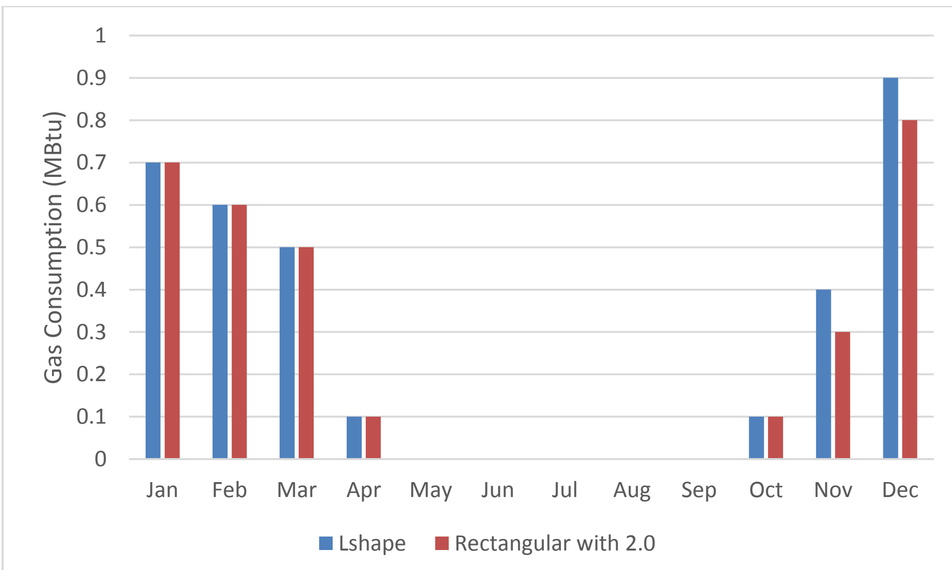


Figure 6-10. Elec and Gas consumption difference result after 10% variation to infiltration rate in 'L-shape' and 'Rectangular with aspect ratio of 2.0'

6.4. Comparison of Bayesian approach

6.4.1. Introduction

In our study, to eliminate the limitations from a bottleneck in the Bayesian approach, the building emulator and sampling was developed. Pavlak's study (Pavlak, 2014) utilizes an RC-model to alleviate the CPU burden in Bayesian approach. According to personal correspondence with the author, 500,000 sampling number was used in sampling process for 5 or 6 parameter identification problem which equates 42 hours of compute time. This indicates heavy CPU burden was needed during the process. To verify if our approach actually alleviates the CPU burden of a traditional Bayesian approach, a comparison was done.

Figure 6-11 shows the procedure for a traditional Bayesian approach and the proposed approach. As shown in the figure, the traditional approach starts with unknown parameters and the observation noise. Then by using these inputs, actual building simulation process is built, which then samples for the posterior distribution. Our approach utilizes building emulator's process prior to sampling process. As stated in previous sections, our approach utilizes 1,000 training points which means 1,000 simulation runs prior to sampling process. During the building emulator process, hyperparameters are optimized and identified. By using pre-determined hyperparameters, we then utilize the Gaussian process emulator to sample the posterior distribution of these unknown parameters.

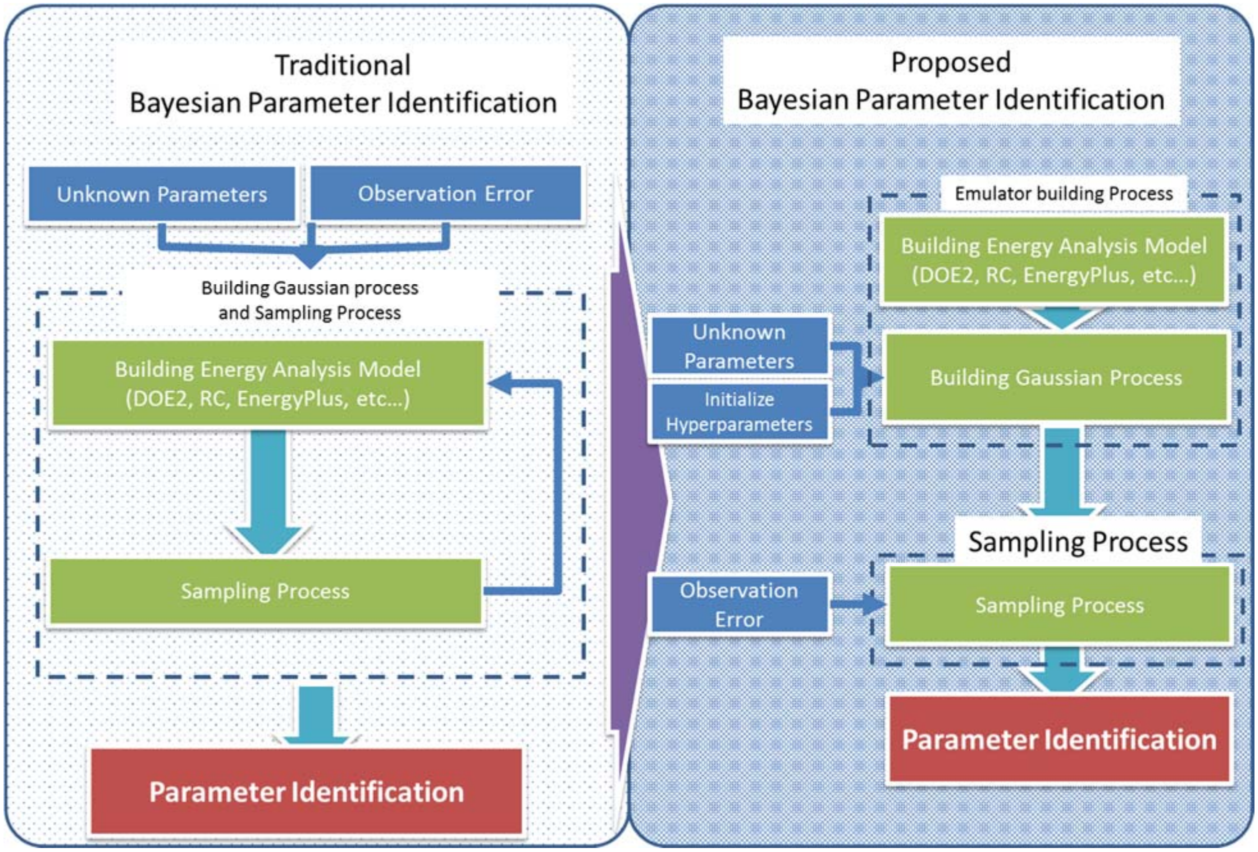


Figure 6-11. Comparison between traditional approach and proposed Bayesian approach

6.4.2. Literature review

Gibbs's (Gibbs, 1997) study mentioned that the performance of Bayesian methods crucially depends on the capability of finding good models for the data. He mentioned that if the number of hyperparameters, as well as the amount of data, increases significantly and that the cost of evaluating the derivatives of the covariance matrix becomes extremely expensive with the cost of inversion. As hyperparameters can be increased by the number of data dimensions (multivariate), this means that the computing burden will increase significantly when the number of unknown variables increases.

Snelson (Snelson, 2007) claimed that the Gaussian process is nonparametric, meaning that the complexity of the entire model grows while data points increases. In his study, the training cost for a Gaussian process has $O(N^3)$ complexity, where N represents the number of training data points. As indicated, the complexity of models grows when more data is received. This is due to an inversion of N X N covariance matrix since it takes a considerable amount of time to compute the inversion of a large matrix. In his study, he developed new techniques to reduce this complexity to $O(NM^2)$, where M is a user chosen number smaller than N. While a standard Gaussian process assumes a uniform noise variance, he used sparse approximation to relieve this assumption. By using sparse covariance function with modifications, he was able to model input dependent noise instead of uniform noise.

Higdon (Higdon, et al., 2004) stated that the limitation on simulation run, which is called training data, contributes greatly to uncertainty in the Bayesian analysis. Basically, as the dimensions of x (observable inputs) and t (calibration inputs) increases, the limited evaluations of $\eta(x, t)$ will increase the uncertainty of the procedure. In one case study, he reduced the number of particles from 10,000 to 8,000 and this reduced number of particles led to a much noisier simulation as compared to a high fidelity simulation. Simulator function $\eta(\cdot,\cdot)$ becomes highly difficult as the dimensionality of the input space increases because the limited number of runs must cover a high dimensional space.

6.4.3. Methodology

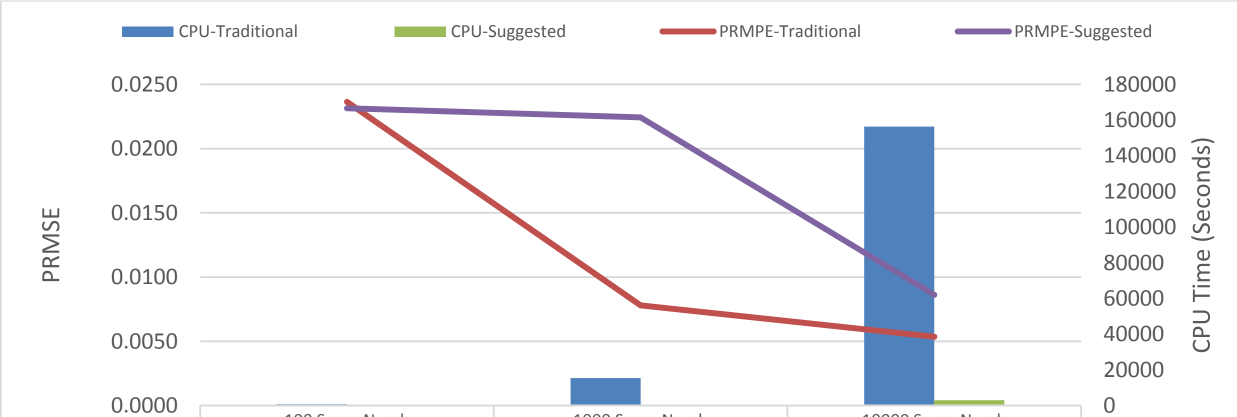
In Higdon's study (Higdon, et al., 2004), there are several ways to pursue a Bayesian approach. One way is to utilize the training data to derive the posterior distribution. In this study, we have utilized training data, which was derived from actual computer simulation (DOE2.2E), to sample the posterior distribution. By comparison, traditional Bayesian approach will follow 'Unlimited simulation runs' that was mentioned in Higdon's study. Basically, this approach will utilize actual computer simulation's result to derive posterior distribution of these unknown parameters. In comparison case study, this study have used 100, 1,000, and 10,000 sampling number to examine the difference between these approaches.

6.4.4. *Results*

Figure 6-12 shows the results for CPU time for a different number of sampling. And Table 6-6 shows the data for the comparison result. As shown in the figure, the accuracy of the traditional approach is actually less than that of proposed approach. Less accuracy is occurred because of actual accuracy of an emulator. Despite of emulator's performance of predicting actual energy consumption, the error occurs since it is mimicking the behavior of actual simulation. However, after 1,000 sampling number, the gap between the traditional approach and proposed approach reduces as the posterior converges to the target value that we are trying to find. This indicate that the accuracy of finding target value can be much less significant if the approach utilizes higher number of sampling. However, as shown in the table, the cpu time of these approaches are significantly different due to its nature of calculation. Since traditional Bayesian approach utilizes actual simulation data, it is over 50 times bigger when performing 10,000 sampling number cases. This means that it is not necessary to utilize actual simulation data if the emulator carries fine training sets.

Table 6-6. Comparison result for traditional Bayesian approach and proposed Bayesian approach

	100 Sampling Number	1000 Sampling Number	10000 Sampling Number
CPU-Traditional (seconds)	771.482	15424.98	156364.1
CPU- proposed (Seconds)	32.836	326.3	3014.8
PRMPE-Traditional	0.0236	0.0078	0.0054
PRMPE- proposed	0.0231	0.0224	0.0086



	100 Samp Numb	1000 Samp Numb	10000 Samp Numb
CPU-Traditional	771.482	15424.983	156364.091
CPU-Suggested	32.836	326.3	3014.8
PRMPE-Traditional	0.0236	0.0078	0.0054
PRMPE-Suggested	0.0231	0.0224	0.0086

Figure 6-12. CPU time comparison of traditional Bayesian approach

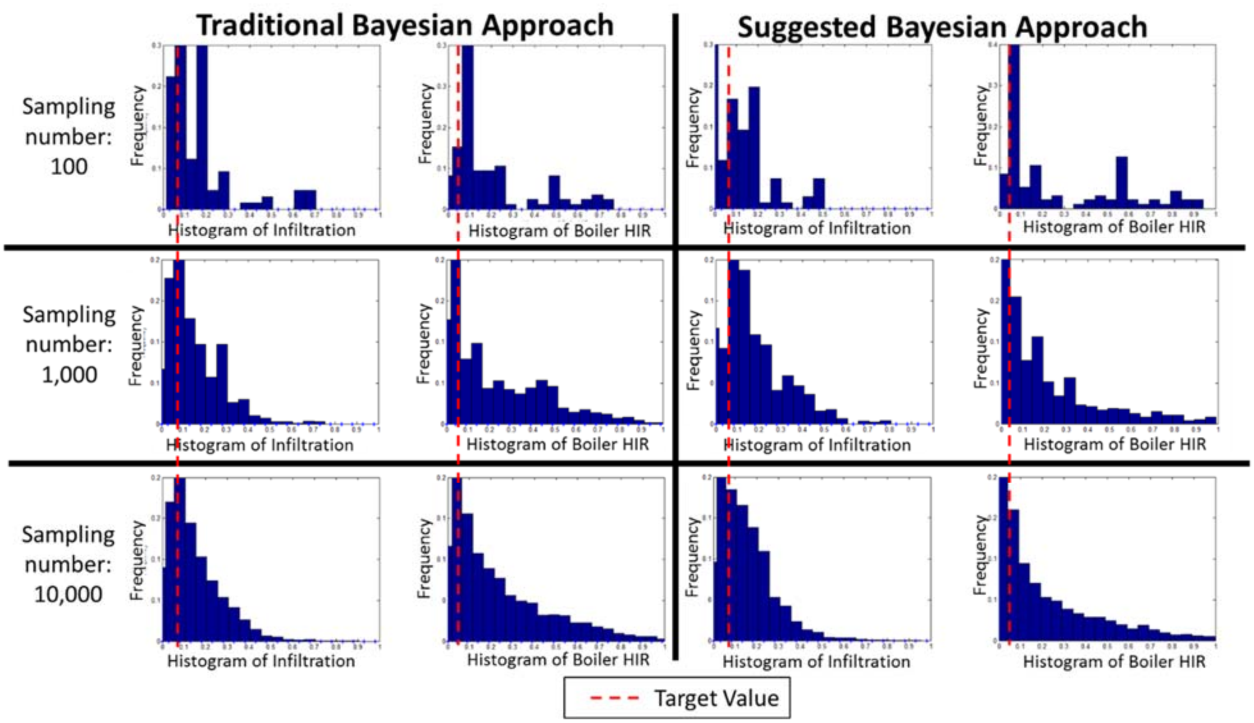


Figure 6-13. Posterior distribution for 2 unknown parameter identification of traditional and proposed Bayesian approach

Figure 6-13 shows the comparison of posterior distribution of these two approaches. As shown in the figure, posterior distribution of both approaches tends to get smoother while the sampling number increases. This indicates that the posterior result tends to be similar since these approaches use the data-likelihood calculation based on same nature of getting one from actual simulation and other from emulator.

6.5. Summary and Conclusions

In this chapter, this study has performed studies about identification of ECMs, comparison to other regression analyses, comparison with the simplified approach, comparison with traditional Bayesian approach.

In identification of ECMs, this study expanded the proposed Bayesian approach to identify potential ECMs. Before going into actual ECMs evaluations, this study has tested the validity of adding additional variables to original Gaussian process emulators. And using four different scenarios, this study has identified optimized energy cost related ECMs. In the result, equipment density had the highest savings due to its high energy saving but low initial costs. And this study has verified that our proposed approach can be expanded to other application such as exploration of ECMs.

Comparison study with other regression models were performed. Four different models of linear regression were compared to Gaussian process model. The result showed that despite of higher demand of computing, Gaussian process had the highest and trustable result on residential building. However, there still relies numerous other regression methods. To validate our study, future study comparing to other methods is needed.

Another study that executes energy model creation with simplified geometry was performed to compare with the model with complex geometry. 4 models were selected, 'L-shape', ‘square’, ‘rectangular with aspect ratio of 1.5’, ‘rectangular with aspect ratio of 2.0’. Parameter identification of ‘L-shape’ was done on other 3 models. Study showed that main energy consumption difference was derived from exterior component’s area. ‘rectangular with aspect ratio of 2.0’ had the closest exterior area as ‘L-shape’ and the result showed similar result. Our parameter identification approach was able to reach nearest value of ‘L-shape’. This study determined the possibility of using our approach with simpler geometry model to identify other complex building models.

Lastly, the comparison study of traditional Bayesian approach was performed. To verify the effectiveness of our approach to traditional Bayesian approach, this study has built traditional Bayesian approach with using actual building energy simulation model. The study result showed that our proposed approach had less accuracy but CPU burden was reduced significantly.

Despite of these studies, there are a lot of applications that the proposed approach can handle. Therefore, it will require further future studies on developing applications to be used in building energy analysis. One of the examples will be combining the simplified model approach with current model to develop a generalized emulator model which can predict energy consumptions and identify unknown parameters with higher accuracy.

CHAPTER 7: Summary and Future Work

7.1. Summary

The purpose of this study is to validate the possibility of utilizing the Bayesian approach to be used in parameter identification problems. Since the traditional Bayesian approach contains drawbacks of requiring higher CPU burden, this study has developed a new framework to alleviate the problem. This study has validated the proposed Bayesian approach by creating an emulator with training points of 1,000. And determined that making pre assumption on the observation noise will help on reducing CPU burdens. Various case studies have been carried to determine the observation noise, size of the Gaussian process, number of sampling. These values were predetermined to be used in the Bayesian framework.

After validation on these pre-selected values for the Bayesian approach, case studies on actual buildings were carried out to determine if the proposed approach succeeds in unknown parameter identification problems in actual situations. Three buildings (two commercial buildings and one residential building) were chosen to participate in these case studies. The results showed that the proposed approach succeeded in parameter identification of physical parameters for buildings. However, since buildings carry both physical parameters and behavioral parameters, there still persists the problem of identification of behavioral parameters as well.

To extend the applicability of utilizing the proposed approach, this study has created a Bayesian method to identify energy conservative measures. The developed approach succeeded in identifying these ECMs with uncertainties. Accounting four different scenarios, ECMs were identified after 1,000 posterior distribution with uncertainty in both physical value and initial cost to these ECMs. Additionally, the ECMs with budget limits were successfully identified.

After validation on the ECMs identification problem, three analyses were performed. The first study was to verify the accuracy and the CPU time when compared with other linear regression methodologies. Four linear models were chosen to be compared with the Gaussian process model. The study has identified that in spite of higher CPU burden, the proposed approach had the most accurate result for the emulator as well as sampling posterior distributions. The higher computing burden can be tolerated since the difference between these models was not significant and the accuracy was higher than the linear models. Even compared with the traditional Bayesian approach, which utilizes actual simulation, the proposed approach had the least accuracy but also had significantly less CPU burden since the proposed approach utilized an emulator.

To test the possibility of extending this approach to simplified analysis, this study has created an energy model with simple geometry to replace a model with complex geometry. The result showed the possibility of utilizing the simple model to be used in the proposed approach to identify the parameters within a model with complex geometry. However, the simplified model required further analysis since there are many parameters involved in the energy consumption differences for a simple model and a complex model.

7.2. Future work

Over the course of concluding the research described in this thesis, a number of prospects have risen to expand the scope of the models to be developed and assessed. Some of the recommendations for future work include:

1. This study performed sensitivity analysis for a Bayesian framework with unknown parameter identification for a RC model and a DOE2 model. The Bayesian framework

resulted in a positive outcome for both models. A follow-up approach can benefit from verification in using other whole building simulation tools (such as Trnsys, EnergyPlus, etc.) in order to achieve generality. ,. In the future, this study will perform sensitivity analysis on a whole-building energy simulation tool such as EnergyPlus and validate the result to be used in the actual building cases.

2. To expand the parameter identification to larger sets of unknowns. In this study, we only considered up to 11 building parameters in the identification process. However, most of the detailed simulation tools require several hundred input parameters. Future work may consider higher number of dimensional problem and this study needs to perform further research on high-dimensional parameter identification problems using the proposed approach. Additionally, in this study, the proposed Bayesian approach only used 24 dimensional reference data (12 monthly electricity usage, 12 monthly gas usage). However in reality, the reference data can be composed of larger data sets. For instance, if a building has 1,000 electric meters and we want to perform the parameter identification using all of the data, which increases the reference data larger than 24 dimensions. To accommodate the larger dimensional reference data, this study will perform the proposed Bayesian approach on larger reference data dimensional problems.

3. The observation noise is crucial component in the posterior probability calculation in the Bayesian approach. Although sensitivity analysis on the observation noise was performed in this study, there still relies an uncertainty in the observation noise selection. There are many different energy analysis tools and various building types and sizes. Therefore, the future work will include more thorough study of the observation noise and perform a study on developing an algorithm that will result in the appropriate observation noise. This may eliminate the need of the manual process of selecting the observation noise for any buildings.

4. To verify the difference of the approach between the optimization methods (such as

PSO, GA) and the Bayesian approach, this study will perform the comparison study regarding these approaches.

5. As mentioned earlier, the simplified model approach can be applied to further reduce CPU time as well as effort to model the building. To validate this approach, more case studies are recommended. One example of future work will be suggested to consider building geometry as a possible parameter to identify.

Bibliography

Alcala Rafael [et al.] A genetic rule weighting and selection process for fuzzy control of heating, ventilating and air conditioning systems [Journal]. - [s.l.] : Engineering Applications of Artificial Intelligence, 2005. - 3 : Vol. 18. - pp. 279 - 296.

Alcala Rafael [et al.] Fuzzy Control of HVAC Systems Optimized by Genetic Algorithms [Journal]. - [s.l.] : Applied Intelligence, 2003. - Vol. 18. - pp. 155 - 177.

ASHRAE ASHRAE Guideline 14-2002: measurement of energy and demand savings [Book]. - Atlanta, GA : American Society of Heating, Refrigerating and Air-conditioning Engineers, 2002.

Assareh E., Behrang M.A. and Ghanbarzadeh A. Application of PSO (particle swarm optimization) and GA (genetic algorithm) techniques on demand estimation of oil in Iran [Journal]. - [s.l.] : Energy, 2010. - 12 : Vol. 35. - pp. 5223 - 5229.

Boujelben A.H. Developing probabilistic and optimized approaches to calibrate residential building energy models based on measured data [Journal]. - [s.l.] : Tunisia Polytechnic School, 2012.

Caldas L. G. and Norford L. K. Genetic Algorithms for Optimization of Building Envelopes and the Design and Control of HVAC Systmes [Journal]. - [s.l.] : Journal of Solar Energy Engineering, 2003. - Vol. 125. - pp. 343 - 351.

Chow T.T. [et al.] Global optimization of absorption chiller system by genetic algorithm and neural network [Journal]. - [s.l.] : Energy and Buildings, 2002. - 1 : Vol. 34. - pp. 103 - 109.

Coehlo J.P., de Moura Oliveira P.B. and Boaventura Cunha J. Greenhouse air temperature predictive control using the particle swarm optimisation algorithm [Journal]. - [s.l.] : Computers and Electronics in Agriculture, 2005. - 3 : Vol. 49. - pp. 330 - 344.

CPU CPU List [Online] // PassMark Software. - PassMark Software Pty Ltd. - http://www.cpubenchmark.net/cpu_list.php.

Currin Carla [et al.] Bayesian Prediction of Deterministic Functions, With Applications to the Design and Analysis of Computer Experiments [Journal]. - [s.l.] : Journal of the American Statistical Association, 1991. - 416 : Vol. 86. - pp. p. 953 - 963.

DOE2 [Online] // DOE2.com Home Page. - http://doe2.com/.

Eisenhower Bryan [et al.] A methodology for meta-model based optimization in builiding energy models [Journal]. - [s.l.] : Energy and Buildings, 2012. - Vol. 47. - pp. 292-301.

Eisenhower Bryan [et al.] Uncertainty and Sensitivity Decomposition of Building Energy Models [Journal]. - [s.l.] : Journal of Building Performance Simulation, 2011. - 42 : Vol. 11.

Ekici Betul Bektas and Aksoy U. Teoman Prediction of building energy consumption by using artificial neural networks [Journal]. - [s.l.] : Advances in Engineering Software, 2009. - 5 : Vol. 40. - pp. 356 - 362.

EPA Greenhouse Gas Equivalencies Calculator [Online] // U.S. Environmental Protection Agency. - 4 2014. - http://www.epa.gov/cleanenergy/energy-resources/calculator.html#results.

Feller William An Introduction to Probability Theory and Its Applications [Book]. - New York : Wiley, 1971.

Fricker Thomas E., Oakley Jeremy E. and Urban Nathan M. Multivariate Gaussian Process Emulators with Nonseparable Covariance Structures [Report]. - [s.l.] : Technical report managing uncertainty in complex models (MUCM), 2010.

Gelman Andrew [et al.] Bayesian Data Analysis [Book]. - [s.l.] : Chapman and Hall/CRC, 2013. - 3rd.

Gibbs Mark N Bayesian Gaussian Processes for Regression and Classification [Book]. -

[s.l.] : Department of Physics, University of Cambridge, 1997. - Vol. PhD Thesis.

Heo Y., Augenbroe G. and Choudhary R. Risk analysis of energy-efficiency projects based on Bayesian calibration of building energy models [Journal]. - [s.l.] : Proceedings of Building Simulation 2011, 2011. - pp. 2579 - 2586.

Heo Y., Choudhary R. and Augenbroe G.A. Calibration of building energy models for retrofit analysis under uncertainty [Journal]. - [s.l.] : Energy and Buildings, 2012. - Vol. 47. - pp. 550 - 560.

Heo Yeonsook Bayesian Calibration of Building Energy Models for Energy Retrofit Decision-Making Under Uncertainty [Book] = PhD Thesis. - [s.l.] : College of Architecture, Georgia Institute of Technology, 2011.

Higdon Dave [et al.] Combining Field Data and Computer Simulations for Calibration and Prediction [Journal]. - [s.l.] : SIAM J. Sci. Comput., 2004. - 2 : Vol. 26. - pp. 448-466.

Hong Tianzhen [et al.] Comparing Computer Run Time of Building Simulation Programs [Journal]. - [s.l.] : Build Simul, 2008. - 3 : Vol. 1. - pp. 210-213.

Hopfield J.J. Neural networks and physical systems with emergent collective computational abilities [Journal]. - [s.l.] : National Acad Sciences, 1982. - 8 : Vol. 79. - pp. 2554 - 2558.

IPMVP International Performance Measurement & Verification Protocol: concepts and options for determining energy and water savings [Book]. - [s.l.] : Efficiency Valuation Organization, 2010. - Vol. I.

Jensen Kasper, Toftum Jorn and Friis-Hansen Peter A Bayesian Network approach to the evaluation of building design and its consequences for employee performance and operational costs [Journal]. - [s.l.] : Building and Environment, 2009. - 3 : Vol. 44. - pp. 456 - 462.

Kennedy C. and O'Hagan Anthony Bayesian calibration of computer models [Journal]. - [s.l.] : Journal of the Royal Statistical Society: Series B , 2001. - 3 : Vol. 63. - pp. 425 - 464.

Kennedy James and Eberhart Russell Particle Swarm Optmization [Conference]. - [s.l.] : IEEE international conference on Neural Networks, 1995. - Vol. 4. - pp. 1942 - 1948.

Kneifel Joshua Life-cycle Carbon and Cost Analysis of Energy Efficiency Measures in New Commercial Buidings [Journal]. - [s.l.] : Energy and Buildings, 2009. - III : Vol. 42. - pp. p. 333 - 340.

Koopmans Tjalling C. Identification Problems in Economic Model Construction [Journal] // Studies in Econometric Method. - New York : Wiley, 1953.

Krarti Moncef Energy Audit of Building Systems: An Engineering Approach [Book]. - [s.l.] : CRC Press, 2010. - 2nd.

Kusiak Andrew and Xu Guanglin Modeling and optimization of HVAC systems using a dynamic neural network [Journal]. - [s.l.] : Energy, 2012. - 1 : Vol. 42. - pp. 241 - 250.

Lauret P., Boland J. and Ridley B. Bayesian statistical analysis applied to solar radiation modeling [Journal]. - [s.l.] : Renewable Energy, 2013. - pp. 124 - 127.

Lavigne Karine Assisted Calibration in Building Simulation-Algorithm Description and Case Studies [Conference] // Eleventh International IBPSA Conference. - Glasgow, Scotland : Building Simulation 2009, 2009. - pp. p. 1498 - 1505.

Lecca Patrizio [et al.] The Added Value from a General Equilibrium Analysis of Increased Efficiency in Household Energy Use [Journal]. - [s.l.] : Ecological Economics, 2014. - Vol. 100. - pp. p. 51 - 62.

Li Xuemei [et al.] Particle Swarm Optimization-based LS-SVM for Building Cooling Load Prediction [Journal]. - [s.l.] : Journal of Computers, 2010. - 4 : Vol. 5. - pp. 614 - 621.

Lilley Matthew and Frean Marcus Neural networks: a replacement for Gaussian processes? [Journal]. - [s.l.] : Springer, 2005. - Vol. 3578. - pp. p. 195 -202.

Liu Li, Liu Wenxin and Cartes David A. Particle swarm optimization-based parameter identification applied to permanent magnet synchronous motors [Journal]. - [s.l.] : Engineering Appications of Artificial Intelligence, 2008. - 7 : Vol. 21. - pp. 1092 - 1100.

Lu Lu [et al.] HVAC system optimization—in-building section [Journal]. - [s.l.] : Energy and Buildings, 2005. - 1 : Vol. 37. - pp. 11 - 22.

MacKay David J. Bayesian Interpolation [Journal]. - [s.l.] : Neural Computation, 1992. - 3 : Vol. 4. - pp. p. 415 - 447.

MacKay David J. C. Introduction to Gaussian Processes [Book]. - [s.l.] : Springer-Verlag, 1998.

Mackay David J.C. Gaussian Processes: A Replacement for Supervised Neural Networks? [Journal] // NIPS Tutorial. - [s.l.] : Cambridge University, 1997.

Manfren Massimiliano, Aste Niccolo and Moshksar Reza Calibration and uncertainty analysis for computer models – A meta-model based approach for integrated building energy simulation [Journal]. - [s.l.] : Applied Energy, 2013. - Vol. 103. - pp. 627 - 641.

Matsumoto Takashi [et al.] Reconstruction and prediction of nonlinear dynamical systems: a hierarchical Bayes approach with neural nets [Journal]. - [s.l.] : 1999 IEEE International Conference on Acoustics, Speech, and Signal Processing, 1999. - Vol. 2. - pp. 1061 - 1064.

McCulloch Warren S. and Pitts Walter A Logical Calculus of the Ideas Immanent in Nervous Activity [Journal]. - [s.l.] : Bulletin of Mathematical Biophysics, 1943. - Vol. 5. - pp. p. 115 - 133.

Mechaqrane A. and Zouak M. A comparison of linear and neural network ARX models applied to a prediction of the indoor temperature of a building [Journal]. - [s.l.] : Neural Computing & Applications, Springer-Verlag London Limited 2004, 2004. - pp. 32 - 37.

MGE Residential Natural Gas Rates [Online] // Madison Gas and Electric. - Jan 1, 2013. - http://www.mge.com/customer-service/home/gas-rates-res/.

Mossolly M., Ghali K. and Ghaddar N. Optimal control strategy for a multi-zone air conditioning system using a genetic algorithm [Journal]. - [s.l.] : Energy, 2009. - 1 : Vol. 34. - pp. 58 - 66.

Najafi Massieh [et al.] Application of machine learning in the fault diagnostics of air handling units [Journal]. - [s.l.] : Applied Energy, 2012. - Vol. 96. - pp. 347 - 358.

Najafi Massieh [et al.] Modeling and Measurement Constraints in Fault Diagnostics for HVAC Systems [Report]. - [s.l.] : Lawrence Berkeley National Laboratory, 2010. - LBNL-3903E.

Neal Radford M. Bayesian Learning for Neural Networks [Book]. - New York : Spring Verlag, 1996.

Nordhaus William A Question of Balance-Weighing the Options on Global Warming Policies [Article]. - [s.l.] : Yale University Press, 2008.

Novak V., Perfilieva I. and Mockor J. Mathematical principles of fuzzy logic [Journal]. - [s.l.] : Dodrecht: Kluwer Academic, 1999.

Office of Energy Efficiency & Renewable Energy Commercial Reference Buildings [Online] // Office of Energy Efficiency & Renewable Energy. - U.S. Department of Energy. - http://energy.gov/eere/buildings/commercial-reference-buildings.

O'neil Patrick J., Crawley Drury B. and Schliesing J. Steven Using Regression Equations to Determine the Relative Importance of Inputs to Energy Simulation Tools [Conference] // Proceedings of the Building Simulation. - Australia : [s.n.], 1991. - pp. p. 283 -

289.

Park Benjamin Thermal analysis of integrated radiant slab heating and cooling systems [Report] : Master Thesis. - [s.l.] : Univ of Colorado at Boulder, 2012.

Pavlak Gregory S. Building-to-Grid Integration through Commercial Building Portfolios Participating in Energy and Frequency Regulation Markets [Book]. - [s.l.] : Univ of Colorado at Boulder, 2014. - PhD Thesis.

Rallapalli Hema Sree A Comparison of EnergyPlus and eQUEST Whole Building Energy Simulation Results for a Medium Sized Office Building [Report] : Master of Science Thesis. - [s.l.] : Arizona State University, 2010.

Rasmussen Carl Edward and William Christopher K. I. Gaussian Processes for Machine Learning [Book]. - [s.l.] : MIT Press, 2006. - 2nd. - ISBN-10 0-262-18253-X, ISBN-13 978-0-262-18253-9.

Rumelhart David E. and McClelland Jame L. Parallel distributed processing: explorations in the microstructure of cognition [Book]. - [s.l.] : MIT Press, 1986.

Rumelhart David E., Hinton Geoffrey E. and Williams Ronald J. Learning Representations by Back-propagating Errors [Journal]. - [s.l.] : Nature, 1986. - Vol. 323. - pp. p. 533 - 536.

Sacks Jerome [et al.] Design and Anlaysis of Computer Experiments [Journal]. - [s.l.] : Statistical Science, 1989. - 4 : Vol. 4. - pp. p. 409 - 435.

Signor Regis, Westphal Fernando Simon and Lamberts Roberto Regression Analysis of Electric Energy Consumption and Architectural Variables of Conditioned Commercial Buildings in 14 Brazilian Cities [Conference] // Proceedings of the Building Simulation. - 2001. - pp. p. 1373 - 1379.

Snelson Ed Tutorial: Gaussian process models for machine learning [Report]. - [s.l.] : Gatsby Computational Neuroscience Unit, UCL, 2006.

Snelson Edward Lloyd Flexible and efficient Gaussian process models for machine learning [Book]. - [s.l.] : Gatsby Computational Neuroscience Unit, University College London, 2007. - Vol. PhD Thesis.

Stankovic Sinisa [et al.] Evaluation of Energy Efficiency Measures applied in Public Buildings (Schools & Hostpitals) in Serbia [Journal]. - [s.l.] : SPATIUM International Review, 2009. - pp. p. 1 - 8.

Steyvers Mark Computational Statistics with Matlab [Book]. - 2011.

Sustainable Buildings & Climate Initiative Buildings and Climate Change [Report]. - Paris : United Nations Environment Programme, 2009.

Sykes Alan O. An Introduction to Regression Analysis [Report].

Tarlow Daniel [et al.] Automatically calibrating a probabilistic graphical model of building energy consumption [Conference] // Eleventh International IBPSA Conference. - [s.l.] : Building Simulation, 2009. - pp. 904 - 911.

Tse Edison and Weinert Howard L. Structure determination and parameter identification for multivariable stochastic linear systems [Journal]. - [s.l.] : IEEE Transactions on Automatic Control, 1975. - 5 : Vol. 20. - pp. 603 - 613.

U.S. DOE [Online] // U.S. Energy Informaiton Administration. - 4 17, 2014b. - http://www.eia.gov/tools/faqs/faq.cfm?id=74&t=11.

U.S. DOE Electric Power Monthly [Online] // U.S. Energy Information Administration. - September 25, 2014a. - http://www.eia.gov/electricity/monthly/epm_table_grapher.cfm?t=epmt_5_06_b.

Urban Bryan J. and Glicksman Leon R. A Simplified Rapid Energy Model and Interface for Nontechnical Users [Conference] // Buildings Conference X. - [s.l.] : Proceedings of Oak Ridge National Laboratory, 2007. - p. Paper#192.

Valade Rachel Elizabeth Development and Verification of a Simplified Building Energy Model [Book] = M.S Degree thesis. - Atlanta, Georgia : Georgia Institute of Technology, 2009.

Wang Sheng and Xu Xinhua Parameter estimation of internal thermal mass of building dynamic models using genetic algorithm [Journal]. - [s.l.] : Energy Conversion and Management, 2006a. - 13-14 : Vol. 47. - pp. 1927 - 1941.

Wang Shengwei and Jin Sinqiao Model-based optimal control of VAV air-conditioning system using genetic algorithm [Journal]. - [s.l.] : Building and Environment, 2000. - 6 : Vol. 35. - pp. 471 - 487.

Wang Shengwei and Ma Zhenjun Supervisory and Optimal Control of Building HVAC Systems: A Review [Journal]. - [s.l.] : HVAC&R Research, 2008. - 1 : Vol. 14. - pp. 3 - 32.

Wang Shengwei and Xu Xinhua Simplified building model for transient thermal performance estimation using GA-based parameter identification [Journal]. - [s.l.] : International Journal of Thermal Science, 2006b. - 4 : Vol. 45. - pp. 419 - 432.

Wang Tzai-Der, Chuang Shang-Jen and Fyfe Colin Comparing Guassian Processes and Artificial Neural Networks for Forecasting [Conference] // Proc of 9th Joint Conf on Information Sciences. - Taiwan : [s.n.], 2006. - pp. p. 1 - 4.

Westphal Fernando Simon and Lamberts Roberto Regression Analysis of Electric Energy Consumption of Commercial Buildings in Brazil [Conference] // Proceedings of Building Simulation 2007. - 2007. - pp. p. 1543 - 1550.

Whitley Darrell A genetic algorithm tutorial [Journal]. - [s.l.] : Statistics and Computing,

1994. - 2 : Vol. 4. - pp. 65 - 85.

Wikipedia Normal distribution [Online] // Wikipedia. - 10 2014. - 10 2014. - http://en.wikipedia.org/wiki/Normal_distribution.

Xu Xinhua [et al.] A model-based optimal ventilation control strategy of multi-zone VAV air-conditioning systems [Journal]. - [s.l.] : Applied Thermal Engineering, 2009. - 1 : Vol. 29. - pp. 91 - 104.

Xu Xinhua and Wang Shengwei Optimal simplified thermal models of building envelope based on frequency domain regression using genetic algorithm [Journal]. - [s.l.] : Energy and Buildings, 2007. - 5 : Vol. 39. - pp. 525 - 536.

Yan Yongmin [et al.] Adaptive optimal control model for building cooling and heating sources [Journal]. - [s.l.] : Energy and Buildings, 2008. - 8 : Vol. 40. - pp. 1394 - 1401.

Yoon Jong-Ho and Lee Euy-Joon Calibration procedure of energy performance simulation model for a commercial building [Journal]. - [s.l.] : Journal of Solar Energy Engineering, 2003.

Zhang Ling [et al.] A genetic-algorithm-based experimental technique for determining heat transfer coefficient of exterior wall surface [Journal]. - [s.l.] : Applied Thermal Engineering, 2004. - 2-3 : Vol. 24. - pp. 339 - 349.

Zhang Tong [et al.] A new research of identification strategy based on particle swarm optimization and least square [Journal]. - [s.l.] : Frontiers of Electrical and Electronic Engineering in China, 2009. - 3 : Vol. 4. - pp. 313 - 317.

APPENDIX A. (Z-table)

Z Value	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
-4	0.000032	0.00003	0.000029	0.000028	0.000027	0.000026	0.000025	0.000024	0.000023	0.000022
-3.9	0.000048	0.000046	0.000044	0.000042	0.000041	0.000039	0.000037	0.000036	0.000034	0.000033
-3.8	0.000072	0.000069	0.000067	0.000064	0.000062	0.000059	0.000057	0.000054	0.000052	0.00005
-3.7	0.000108	0.000104	0.0001	0.000096	0.000092	0.000088	0.000085	0.000082	0.000078	0.000075
-3.6	0.000159	0.000153	0.000147	0.000142	0.000136	0.000131	0.000126	0.000121	0.000117	0.000112
-3.5	0.000233	0.000224	0.000216	0.000208	0.0002	0.000193	0.000185	0.000178	0.000172	0.000165
-3.4	0.000337	0.000325	0.000313	0.000302	0.000291	0.00028	0.00027	0.00026	0.000251	0.000242
-3.3	0.000483	0.000466	0.00045	0.000434	0.000419	0.000404	0.00039	0.000376	0.000362	0.000349
-3.2	0.000687	0.000664	0.000641	0.000619	0.000598	0.000577	0.000557	0.000538	0.000519	0.000501
-3.1	0.000968	0.000935	0.000904	0.000874	0.000845	0.000816	0.000789	0.000762	0.000736	0.000711
-3	0.00135	0.001306	0.001264	0.001223	0.001183	0.001144	0.001107	0.00107	0.001035	0.001001
-2.9	0.001866	0.001807	0.00175	0.001695	0.001641	0.001589	0.001538	0.001489	0.001441	0.001395
-2.8	0.002555	0.002477	0.002401	0.002327	0.002256	0.002186	0.002118	0.002052	0.001988	0.001926
-2.7	0.003467	0.003364	0.003264	0.003167	0.003072	0.00298	0.00289	0.002803	0.002718	0.002635
-2.6	0.004661	0.004527	0.004396	0.004269	0.004145	0.004025	0.003907	0.003793	0.003681	0.003573
-2.5	0.00621	0.006037	0.005868	0.005703	0.005543	0.005386	0.005234	0.005085	0.00494	0.004799
-2.4	0.008198	0.007976	0.00776	0.007549	0.007344	0.007143	0.006947	0.006756	0.006569	0.006387
-2.3	0.010724	0.010444	0.01017	0.009903	0.009642	0.009387	0.009137	0.008894	0.008656	0.008424
-2.2	0.013903	0.013553	0.013209	0.012874	0.012545	0.012224	0.011911	0.011604	0.011304	0.011011
-2.1	0.017864	0.017429	0.017003	0.016586	0.016177	0.015778	0.015386	0.015003	0.014629	0.014262
-2	0.02275	0.022216	0.021692	0.021178	0.020675	0.020182	0.019699	0.019226	0.018763	0.018309
-1.9	0.028717	0.028067	0.027429	0.026803	0.02619	0.025588	0.024998	0.024419	0.023852	0.023295
-1.8	0.03593	0.035148	0.03438	0.033625	0.032884	0.032157	0.031443	0.030742	0.030054	0.029379
-1.7	0.044565	0.043633	0.042716	0.041815	0.04093	0.040059	0.039204	0.038364	0.037538	0.036727
-1.6	0.054799	0.053699	0.052616	0.051551	0.050503	0.049471	0.048457	0.04746	0.046479	0.045514
-1.5	0.066807	0.065522	0.064255	0.063008	0.06178	0.060571	0.05938	0.058208	0.057053	0.055917
-1.4	0.080757	0.07927	0.077804	0.076359	0.074934	0.073529	0.072145	0.070781	0.069437	0.068112
-1.3	0.0968	0.095098	0.093418	0.091759	0.090123	0.088508	0.086915	0.085343	0.083793	0.082264
-1.2	0.11507	0.113139	0.111232	0.109349	0.107488	0.10565	0.103835	0.102042	0.100273	0.098525
-1.1	0.135666	0.1335	0.131357	0.129238	0.127143	0.125072	0.123024	0.121	0.119	0.117023
-1	0.158655	0.156248	0.153864	0.151505	0.14917	0.146859	0.144572	0.14231	0.140071	0.137857
-0.9	0.18406	0.181411	0.178786	0.176186	0.173609	0.171056	0.168528	0.166023	0.163543	0.161087
-0.8	0.211855	0.20897	0.206108	0.203269	0.200454	0.197663	0.194895	0.19215	0.18943	0.186733
-0.7	0.241964	0.238852	0.235762	0.232695	0.22965	0.226627	0.223627	0.22065	0.217695	0.214764

-0.6	0.274253	0.270931	0.267629	0.264347	0.261086	0.257846	0.254627	0.251429	0.248252	0.245097
-0.5	0.308538	0.305026	0.301532	0.298056	0.294599	0.29116	0.28774	0.284339	0.280957	0.277595
-0.4	0.344578	0.340903	0.337243	0.333598	0.329969	0.326355	0.322758	0.319178	0.315614	0.312067
-0.3	0.382089	0.37828	0.374484	0.3707	0.366928	0.363169	0.359424	0.355691	0.351973	0.348268
-0.2	0.42074	0.416834	0.412936	0.409046	0.405165	0.401294	0.397432	0.39358	0.389739	0.385908
-0.1	0.460172	0.456205	0.452242	0.448283	0.44433	0.440382	0.436441	0.432505	0.428576	0.424655
0	0.5	0.503989	0.507978	0.511966	0.515953	0.519939	0.523922	0.527903	0.531881	0.535856
0.1	0.539828	0.543795	0.547758	0.551717	0.55567	0.559618	0.563559	0.567495	0.571424	0.575345
0.2	0.57926	0.583166	0.587064	0.590954	0.594835	0.598706	0.602568	0.60642	0.610261	0.614092
0.3	0.617911	0.62172	0.625516	0.6293	0.633072	0.636831	0.640576	0.644309	0.648027	0.651732
0.4	0.655422	0.659097	0.662757	0.666402	0.670031	0.673645	0.677242	0.680822	0.684386	0.687933
0.5	0.691462	0.694974	0.698468	0.701944	0.705401	0.70884	0.71226	0.715661	0.719043	0.722405
0.6	0.725747	0.729069	0.732371	0.735653	0.738914	0.742154	0.745373	0.748571	0.751748	0.754903
0.7	0.758036	0.761148	0.764238	0.767305	0.77035	0.773373	0.776373	0.77935	0.782305	0.785236
0.8	0.788145	0.79103	0.793892	0.796731	0.799546	0.802337	0.805105	0.80785	0.81057	0.813267
0.9	0.81594	0.818589	0.821214	0.823814	0.826391	0.828944	0.831472	0.833977	0.836457	0.838913
1	0.841345	0.843752	0.846136	0.848495	0.85083	0.853141	0.855428	0.85769	0.859929	0.862143
1.1	0.864334	0.8665	0.868643	0.870762	0.872857	0.874928	0.876976	0.879	0.881	0.882977
1.2	0.88493	0.886861	0.888768	0.890651	0.892512	0.89435	0.896165	0.897958	0.899727	0.901475
1.3	0.9032	0.904902	0.906582	0.908241	0.909877	0.911492	0.913085	0.914657	0.916207	0.917736
1.4	0.919243	0.92073	0.922196	0.923641	0.925066	0.926471	0.927855	0.929219	0.930563	0.931888
1.5	0.933193	0.934478	0.935745	0.936992	0.93822	0.939429	0.94062	0.941792	0.942947	0.944083
1.6	0.945201	0.946301	0.947384	0.948449	0.949497	0.950529	0.951543	0.95254	0.953521	0.954486
1.7	0.955435	0.956367	0.957284	0.958185	0.95907	0.959941	0.960796	0.961636	0.962462	0.963273
1.8	0.96407	0.964852	0.96562	0.966375	0.967116	0.967843	0.968557	0.969258	0.969946	0.970621
1.9	0.971283	0.971933	0.972571	0.973197	0.97381	0.974412	0.975002	0.975581	0.976148	0.976705
2	0.97725	0.977784	0.978308	0.978822	0.979325	0.979818	0.980301	0.980774	0.981237	0.981691
2.1	0.982136	0.982571	0.982997	0.983414	0.983823	0.984222	0.984614	0.984997	0.985371	0.985738
2.2	0.986097	0.986447	0.986791	0.987126	0.987455	0.987776	0.988089	0.988396	0.988696	0.988989
2.3	0.989276	0.989556	0.98983	0.990097	0.990358	0.990613	0.990863	0.991106	0.991344	0.991576
2.4	0.991802	0.992024	0.99224	0.992451	0.992656	0.992857	0.993053	0.993244	0.993431	0.993613
2.5	0.99379	0.993963	0.994132	0.994297	0.994457	0.994614	0.994766	0.994915	0.99506	0.995201
2.6	0.995339	0.995473	0.995604	0.995731	0.995855	0.995975	0.996093	0.996207	0.996319	0.996427
2.7	0.996533	0.996636	0.996736	0.996833	0.996928	0.99702	0.99711	0.997197	0.997282	0.997365
2.8	0.997445	0.997523	0.997599	0.997673	0.997744	0.997814	0.997882	0.997948	0.998012	0.998074
2.9	0.998134	0.998193	0.99825	0.998305	0.998359	0.998411	0.998462	0.998511	0.998559	0.998605

3	0.99865	0.998694	0.998736	0.998777	0.998817	0.998856	0.998893	0.99893	0.998965	0.998999
3.1	0.999032	0.999065	0.999096	0.999126	0.999155	0.999184	0.999211	0.999238	0.999264	0.999289
3.2	0.999313	0.999336	0.999359	0.999381	0.999402	0.999423	0.999443	0.999462	0.999481	0.999499
3.3	0.999517	0.999534	0.99955	0.999566	0.999581	0.999596	0.99961	0.999624	0.999638	0.999651
3.4	0.999663	0.999675	0.999687	0.999698	0.999709	0.99972	0.99973	0.99974	0.999749	0.999758
3.5	0.999767	0.999776	0.999784	0.999792	0.9998	0.999807	0.999815	0.999822	0.999828	0.999835
3.6	0.999841	0.999847	0.999853	0.999858	0.999864	0.999869	0.999874	0.999879	0.999883	0.999888
3.7	0.999892	0.999896	0.9999	0.999904	0.999908	0.999912	0.999915	0.999918	0.999922	0.999925
3.8	0.999928	0.999931	0.999933	0.999936	0.999938	0.999941	0.999943	0.999946	0.999948	0.99995
3.9	0.999952	0.999954	0.999956	0.999958	0.999959	0.999961	0.999963	0.999964	0.999966	0.999967
4	0.999968	0.99997	0.999971	0.999972	0.999973	0.999974	0.999975	0.999976	0.999977	0.999978