UNCERTAINTY ANALYSIS IN BUILDING ENERGY SIMULATION: A PRACTICAL APPROACH

Adrian Chong[1], Weili Xu[1], and Khee Poh Lam[1]
[1]Center for Building Performance and Diagnostics, Carnegie Mellon University, Pittsburgh, USA

ABSTRACT

Uncertainty analysis in building energy simulation is often carried out with Monte Carlo Analysis. However, there is currently no standard framework for uncertainty modeling in building energy models. In particular, uncertainty quantification is often based on literature and expert judgment with limited attention on the use of measured data. The purpose of this paper is to provide a structured framework for modeling uncertainties in building energy models. A method for selecting probability distributions based on measured data is presented, allowing users to assign probability distributions that better represents actual data distributions. If measured data is not available, the subjective assignment of an appropriate probability distribution would be based on drawings, specifications, literature and past case studies on uncertainty quantification. To facilitate its implementation to building design and retrofit analysis, an application that couples the proposed framework with EnergyPlus is introduced. Demonstrated through a case study, the proposed framework provides acceptable simulation results in comparison with measured gas and electricity consumption.

INTRODUCTION

Building energy simulation has the potential to guide building design by providing a means to quantify and evaluate the energy efficiency of design alternatives. Calibrated simulation is also recommended for establishing benchmark energy consumption when such data is not available or the performance of each energy conservation measure needs to be evaluated individually (EVO, 2012). An accurate model requires users to provide a detailed description of a building's geometry and construction, its associated mechanical systems and the quantification of various internal loads. However, detailed information on building materials, components and systems specification needed for energy simulation are often not available in reality. As a result, the modeler is often forced to set unknown parameters to default values and best guesses. Furthermore, the complexity of building systems often make it necessary to introduce simplifications into the simulation model (De Wit, 2004). These simplifications add to the uncertainties in the simulation outcome. However, uncertainty analysis is still the exception rather than the norm.

Although uncommon the use of uncertainty analysis in building performance simulation is not new. This is because current tools do not facilitate the inclusion of uncertainty analysis in the simulation process. To date, several studies have been conducted on the use of uncertainty analysis in building performance simulation (De Wit and Augenbroe, 2002; Struck and Hensen, 2007; Hopfe and Hensen, 2011; Eisenhower et al., 2012; Lee et al., 2013). This process often involves the quantification of uncertainties in the model's inputs, using information from literature, subjective assessment or expert judgement. Uncertainties in the simulation outcome are then quantified using Monte Carlo Analysis. This process is cumbersome and unsuitable for practical applications since it often involves a manual changing of inputs or the writing of specific programs that needs to be modified each time the programs are used for a new project. Furthermore, Monte Carlo simulation requires running many simulations and generating many output files that can be time-consuming to process.

Modeling uncertainties based on observed data has also received limited attention, with most studies assuming a Gaussian, uniform or triangle distribution which may not be representative of the actual data distribution. Although Bayesian calibration has been shown to be effective, it is currently not suitable for use with building performance simulation programs such as EnergyPlus and ESP-r due to their large number of input parameters (Heo et al., 2012). As a result, the use of uncertainty analysis has not been widespread in practice, even though it offers decision-makers greater assurance in the results generated by simulation tools.

This paper extends current approaches by describing a systematic approach to model uncertainties for practical applications in building performance simulation. A method for assigning a parametric probability distribution based on observed data is presented. If observed data is not available, a knowledge base from past studies and literature is used to guide users on the subjective assignment of an appropriate probability distribution. To facilitate this process, an application that combines uncertainty modeling with the building energy simulation program EnergyPlus is introduced and demonstrated in this paper.

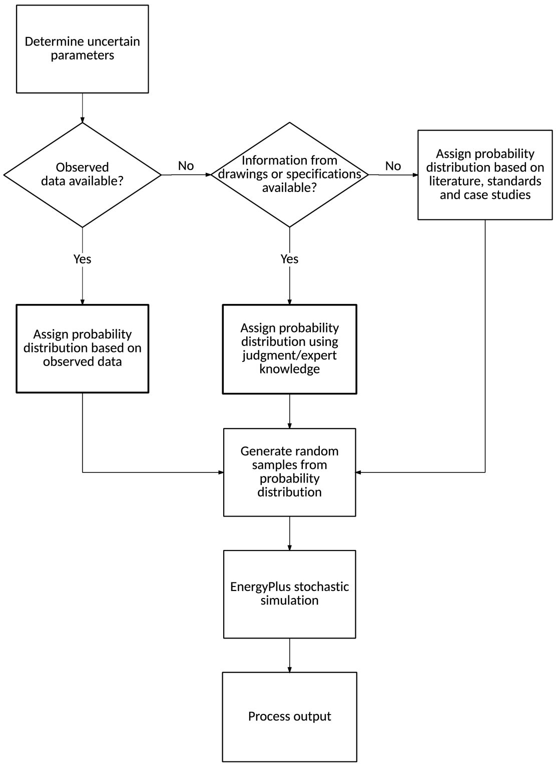

Figure 1 Framework to model uncertainty in building energy simulation.

METHOD

Framework for uncertainty analysis

Figure 1 shows a framework for the quantification of uncertainties in building performance simulation and can be summarized by the following steps:

1. Determine the parameters that are uncertain in the building energy model.
2. For each uncertain parameter,
   (a) First determine if observed data is available. If it is available, fit different parametric probability distributions to the data and select the one that gives the best fit.
   (b) If observed data is not available, determine if a parametric probability distribution can be assigned using subjective assessment of available information such as drawings and specifications.
   (c) If no relevant information from drawings and specifications are available, assign a parametric probability distribution based on current standards, literature and case studies on uncertainty analysis in the field of building performance simulation.
   (d) Using the parametric probability distribution determined by steps 2a to 2c, generate random samples for the uncertain parameter. For this study and in the developed application, Latin hypercube sampling (LHS) is used to sample from the assigned probability distribution. Because of its efficient stratification process, LHS produces results that are more stable than crude Monte Carlo sampling and provides advantages in the analysis of complex models with long run-times (Reddy, 2011).
3. Use the random samples generated for each uncertain parameter as inputs to the building energy model and run the simulations. It should be noted that if $m \times n$ random samples were generated, where n is the number of uncertain parameters and m is the number of samples drawn, there would be m simulation models since each set of n random samples corresponds to the uncertain input parameters in a single simulation model.
4. Process the output files and analyze the results after the simulations are completed.

To facilitate this process, an application that combines uncertainty modeling with whole building energy simulation program EnergyPlus was created. The use of this application will be demonstrated with a case study.

Fitting distributions to data

If observed data is available, the probability distribution should be constructed based on observations instead of using subjective assessment. This can be achieved by fitting different parametric probability distributions to the data and selecting the one that gives the best fit based on some criterion (Sheppard, 2012). Possible criteria include the negative log-likelihood (Equation 1), Bayesian information criterion (BIC) (Equation 2), Akaike information criterion (AIC) (Equation 3) and AIC with a correction for finite sample size (AICc) (Equation 4).

$$-\log \mathcal{L}(\theta|x) = -\log \prod_{i=1}^{n} f(X_i|\theta) \tag{1}$$

$$BIC = -2\log \mathcal{L}(\theta|x) + k\log(n) \tag{2}$$

$$AIC = -2\log \mathcal{L}(\theta|x) + 2k \tag{3}$$

$$AICc = AIC + \frac{2k(k+1)}{n-k-1} \tag{4}$$

θ denotes the parameters of the probability distribution and is determined using maximum likelihood estimation. $X_1, ..., X_n$ denotes the observed data and are assumed to be independent and identically distributed ($i.i.d.$). n is the number of observations and k is the number of parameters to be estimated. As an example, if the probability distribution under consideration is the normal distribution, then $k = 2$ and $\theta = (\mu, \sigma)$ since the normal distribution is defined by two parameters, the mean (μ) and the standard deviation (σ). Given $X_1, ..., X_n$, maximum likelihood estimation maximizes $\mathcal{L}(\theta|x)$ over all possible θ.

For this study, each set of observed data is fitted to 17 continuous probability distributions or 3 discrete probability distributions accordingly. The continuous distributions include the Beta, Birnbaum-Saunders, Exponential, Extreme Value, Gamma, Generalized Extreme Value, Generalized Pareto, Inverse Gaussian, Logistic, Loglogistic, Lognormal, Nakagami, Normal, Rayleigh, Rician, t Location-Scale and Weilbull distribution. The discrete distributions include the Binomial, Negative Binomial and Poisson distribution. The probability distribution that gives the minimum negative log-likelihood, BIC, AIC or AICc is then selected and used to generate inputs to the building energy model.

Create probability distribution

If observed data is not available, the assignment will be based on subjective assessment that is derived from a pool of information such as equipment specifications, nameplates, industry standards, expert judgement, building surveys, etc. However, if both observed data and relevant information from this pool of information are not available, assignment of an appropriate probability distribution would be based on literature and past case studies on uncertainty quantification. This is further illustrated in the case study below.

APPLICATION

Case study

The building analyzed is an actual ten story office building located in Pennsylvania, U.S.A. The building geometry is created using DesignBuilder V3.2 (Figure 2) which is then exported as an EnergyPlus model. EnergyPlus V8.0 is then used to create all other building systems and schedules.

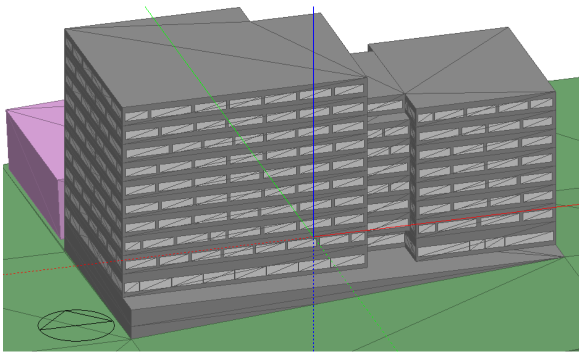

Figure 2 Geometry of office building in DesignBuilder.

The HVAC system is a dual duct system where both warm and cold air are separately ducted and mixed at each terminal unit to achieve the desired temperature. Cooling is supplied with a water cooled chiller while heating is supplied through gas boilers. Data collection for this study took place between 16 March 2013 and 18 December 2013. The weather file used for the simulation is the Actual Meteorological Year (AMY) weather data of the nearest weather station for the period of the data collection (DOE, 2014).

649 parameters were modeled as uncertain in the EnergyPlus model. This is done by setting the value of each parameter to a string that begins with a dollar ($) sign (point 1 in Figure 3). Loading the modified EnergyPlus input data file (IDF) into the application would then list out all random variables (uncertain parameters) on the left panel (points 2 and 3 in Figure 3). Coupling with EnergyPlus is achieved by reading the IDF and storing each EnergyPlus object as key-value pairs in computer memory. These key-value pairs can be modified and exported as a EnergyPlus IDF for simulation. Executing simulations in parallel is achieved by calling the EnergyPlus RunDirMulti.bat file.

Uncertainty quantification based on data

573 of the 649 parameters were assigned probability distributions based on observed data. Figure 4 shows the application's interface when fitting parametric probability distributions to the data. Users should first load a comma-separated (CSV) file containing the data (point 1 in Figure 4). An option to load a single CSV file containing multiple datasets is also provided. The CSV file should be in a format where each column contains the observed values for one uncertain parameter. The first row of the file should also contain the names of the uncertain parameters and match how they are defined in the EnergyPlus IDF.
Added functionality includes the flexibility to select the criterion (negative log-likelihood, BIC, AIC or AICc) for determining the probability distribution that best fits observed data (point 2 in Figure 4), and the option to truncate the distribution to a lower and upper bound (point 3 of Figure 4). The default is to truncate the selected distribution to the minimum and maximum of the observed data. After fitting all parametric probability distributions to the data, details of the distribution that gives the best fit is displayed in the left panel (point 4 in Figure 4). A graph of the top 3 probability distributions is also displayed (point 5 in Figure 4). Lastly, a histogram plot of the samples generated from the selected distribution is shown (point 6 in Figure 4).

Schedules as random variables

Table 1
Modeling schedules as random variables.

Type	Hour	Schedule Value
weekday	$0-1$	random variable 1
	$1-2$	random variable 2
	$\vdots$	$\vdots$
	$23-24$	random variable 24
weekend	$0-1$	random variable 25
	$1-2$	random variable 26
	$\vdots$	$\vdots$
	$23-24$	random variable 48

Schedules form an important part of building energy modeling and allows the user to influence the scheduling of many items (such as occupancy density, lighting and thermostatic controls) (UIUC and LBNL, 2014). To model a building in EnergyPlus users are required to provide a schedule value for each hour of the simulation period. For the case study, schedules were modeled as random variables by defining each schedule to contain a weekday and weekend part with each hour being defined as a random variable (Table 1). This means that the schedule for each item would contain 48 random variables.

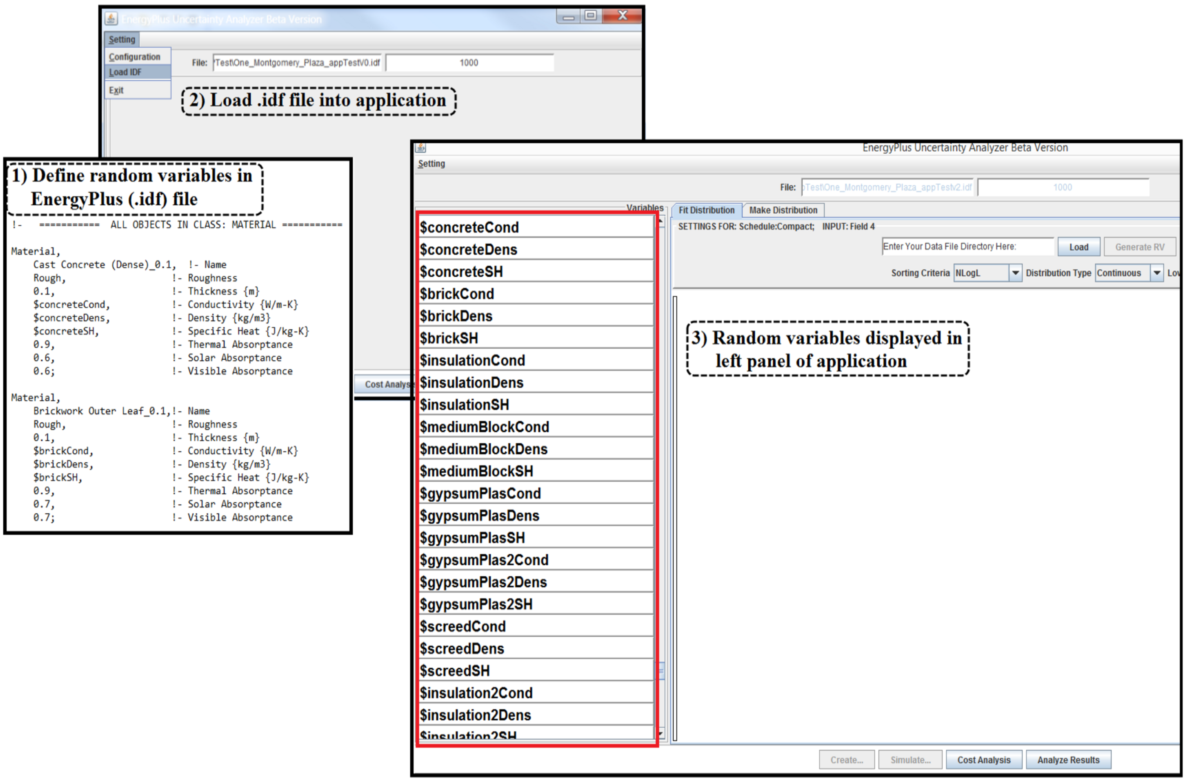

Figure 3 Defining random variables in EnergyPlus IDF and loading it into the application.

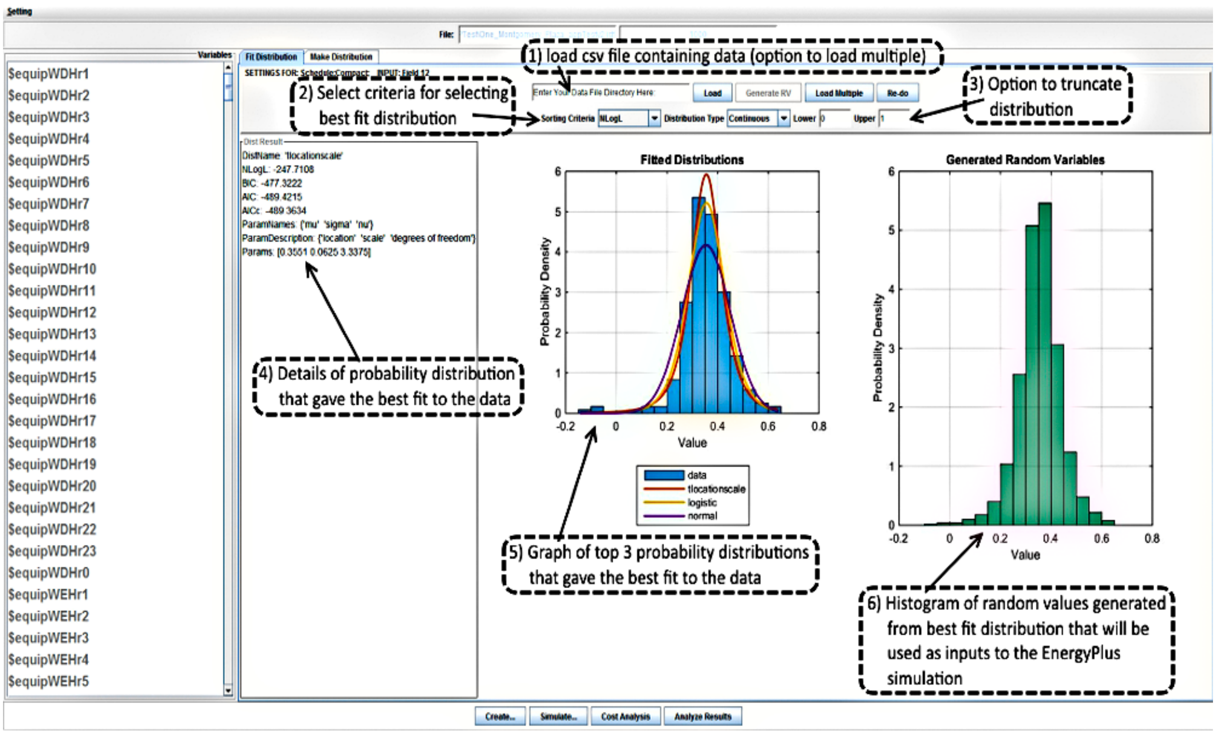

Figure 4 Fit probability distributions to data and using best fit distribution to generate random values.

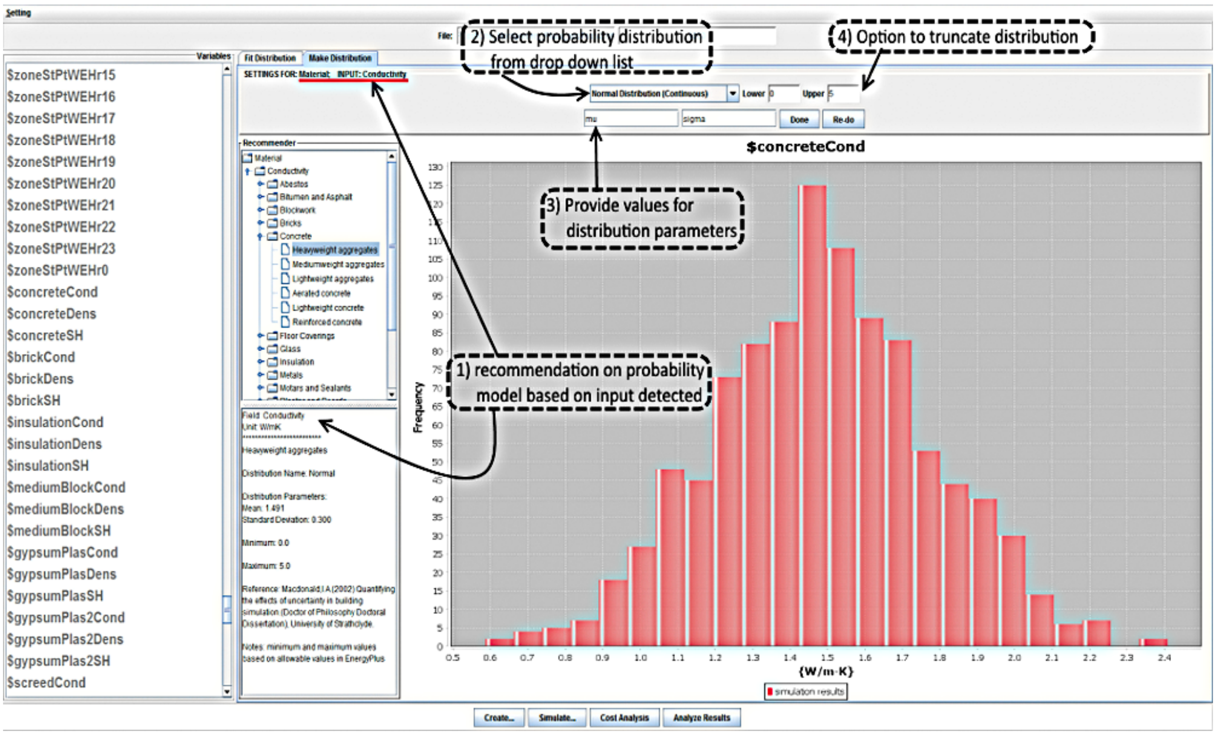

Figure 5 Generate random values from specified probability distribution model.

Uncertainty quantification without data

Uncertain parameters without observed data but whose values can be obtained from specifications and drawings were assigned a uniform distribution with upper and lower bounds that are $\pm 10\%$ of their nominal values respectively. The remaining parameters were modeled based on information from literature, existing standards and past case studies. Uncertainties in thermal properties of materials (conductivity, density and specific heat capacity) were modeled as quantified by Macdonald (2002). Infiltration across the building's exterior surface was modeled as a truncated normal distribution with $\mu = 0.0012\ m^3/s.m^2$, $\sigma = 0.0015\ m^3/s.m^2$ and lower and upper bounds of $0.00011\ m^3/s.m^2$ and $0.0069\ m^3/s.m^2$ respectively (Emmerich and Persily, 2005). These values were obtained by converting the values presented in Emmerich and Persily (2005) from $75Pa$ to $4Pa$, assuming a flow exponent of 0.65. The fraction of radiant heat given off by people is assigned a $\mathcal{U}(0.27, 0.60)$ probability distribution, assuming that the occupants are seated and doing light work (ASHRAE, 2009). The fraction of radiant heat given off by lights is assigned a $\mathcal{U}(0.25, 0.37)$ probability distribution (Chantrasrisalai and Fisher, 2007).

Pairing every uncertain parameter with distribution types from literature is very time consuming. To ease this process, a database pairing various inputs in EnergyPlus to an appropriate probability distribution was created. The database also includes the source of information and any assumptions that was used in the derivation of the recommended probability model. The application reads this database and recommends probability distributions based on the input detected. As an example, the uncertain parameter in the EnergyPlus IDF has been identified as the conductivity of a material (Figure 5). The application searches through the database and displays all probability models for material conductivity in the left panel. Using this application, users are able to easily create probability distributions (points 2 and 3 in Figure 5) from which values would then be sampled from; These values act as inputs to the building energy models. Similar to fitting distributions to data, users are also provided with the option to truncate the distribution (point 4 in Figure 5).

DISCUSSION AND RESULT ANALYSIS

Number of simulations

1000 EnergyPlus simulations were run using the proposed framework and with the help of the application. Suppose that the outputs of the EnergyPlus simulations $X_1, ..., X_n$ are $i.i.d.$ with mean μ and variance σ. The central limit theorem states that $\overline{X} = n^{-1}\sum_{i=1}^{n} X_i$ has a distribution that is approximately Normal with mean μ and variance σ^2/n (Wasserman, 2011). This means that probability statements about $\overline{X}$ can be approximated with a normal distribution. Hence, a normal-based confidence interval can be used to estimate the number of simulations needed in order to achieve reliable results.

Figure 6 shows a plot of a normalized confidence interval ($\frac{95\%\text{ confidence interval}}{\text{standard deviation}}$) against the total number of simulations. It can be observed that after 200-300 simulations, additional simulations have a marginal effect on increasing confidence in the estimation of the population mean. This suggests that the proposed method requires about 200-300 simulations.

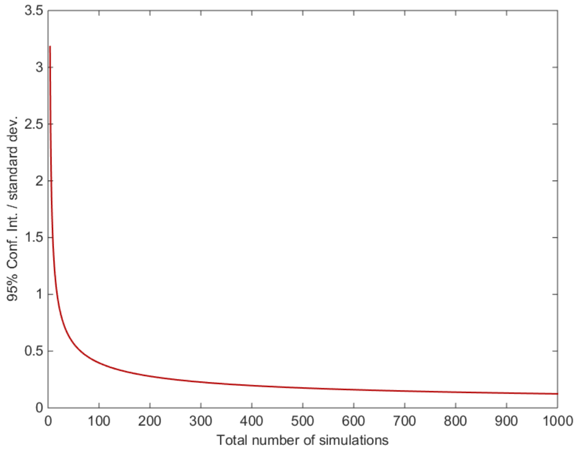

Figure 6 Normalized confidence interval against total number of simulations.

Uncertainty analysis

Data collection for this study took place between 16 March 2013 and 18 December 2013. To allow for comparison, the energy simulations were run over the same period of time. Figure 7 shows a histogram of the predicted total gas energy consumption from the 1000 EnergyPlus simulations normalized by the total building area. From the simulation results, the normalized gas energy consumption has a mean of 51.2 kWh/m^2 and standard deviation of 9.08 kWh/m^2. From Figure 7, it can be observed that the total measured gas energy consumption (46.2 kWh/m^2) falls within one standard deviation of the mean of the predictions. A similar observation can be made for the total electric energy consumption (Figure 8). From the results, the normalized electric energy consumption has a mean of 94.6 kWh/m^2 and a standard deviation of 4.67 kWh/m^2. Measured data (96.7 kWh/m^2) also falls within one standard deviation from the mean of the predictions (Figure 8).

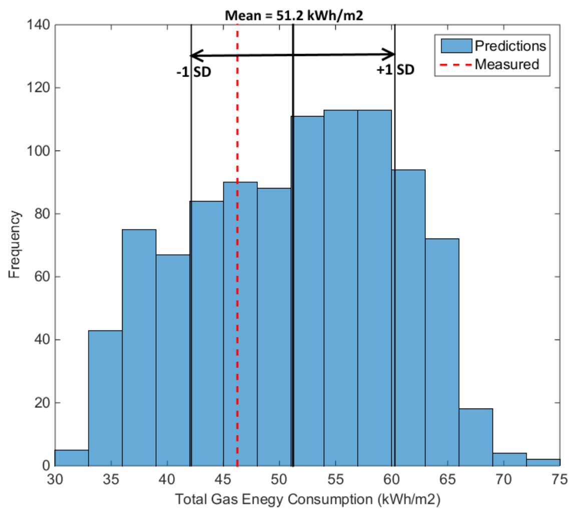

Figure 7 Histogram of predicted total gas energy consumption (March 2013 to December 2013).

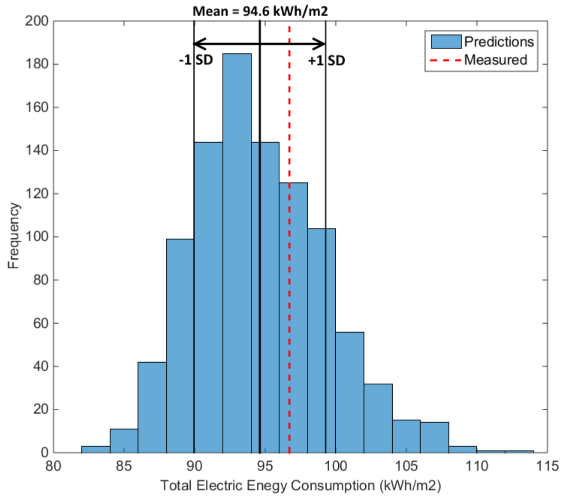

Figure 8 Histogram of predicted total electric energy consumption (March 2013 to December 2013).

To further analyze the simulation results, Figure 9 shows a box plot of the predicted gas energy consumption from the 1000 EnergyPlus simulations at a monthly resolution. The gas energy consumption for each month was normalized by dividing it by the total building area. Using a box plot provides a graphical display of the predictions' symmetry, skewness and spread at a glance. In addition, outliers are also shown. An outlier is defined as one that falls beyond the quartiles (25^{th} and 75^{th} percentile) by 1.5 times the inter-quartile range. It can be observed that measured gas energy consumption falls within the minimum and maximum values predicted by the energy models. This is with the exception of the month of March where the observed value is slightly higher than the largest value predicted by the energy models.

A similar observation can be made from Figure 10, which shows a box plot of the normalized monthly predicted electric energy consumption and how it compares with measured data. It can be seen that the measured electric energy consumption falls within the range predicted by the energy models, with March consumption being slightly above the maximum value and amongst the outliers. By providing probabilistic estimates instead of the usual point estimate with no measure of uncertainty, the proposed framework can contribute to risk quantification. As a result, decision makers would have greater confidence in the models since they are provided with more information when evaluating alternative designs and energy conservation measures.

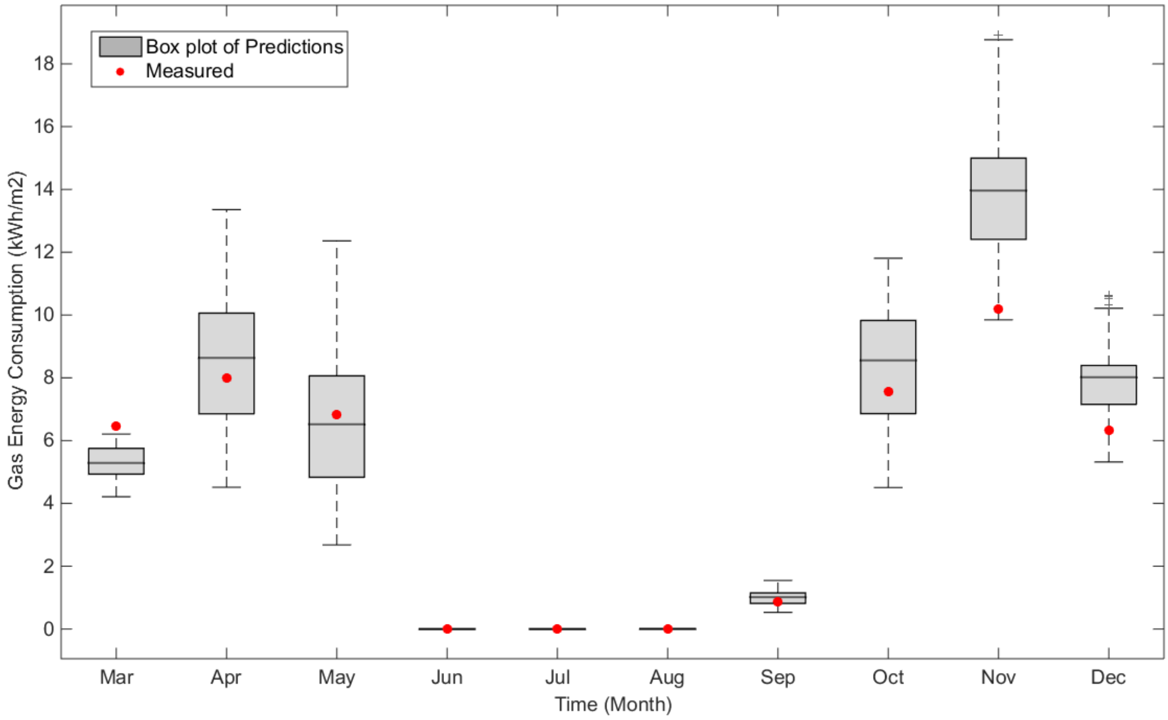

Figure 9 Comparison of model predictions with measured gas consumption.

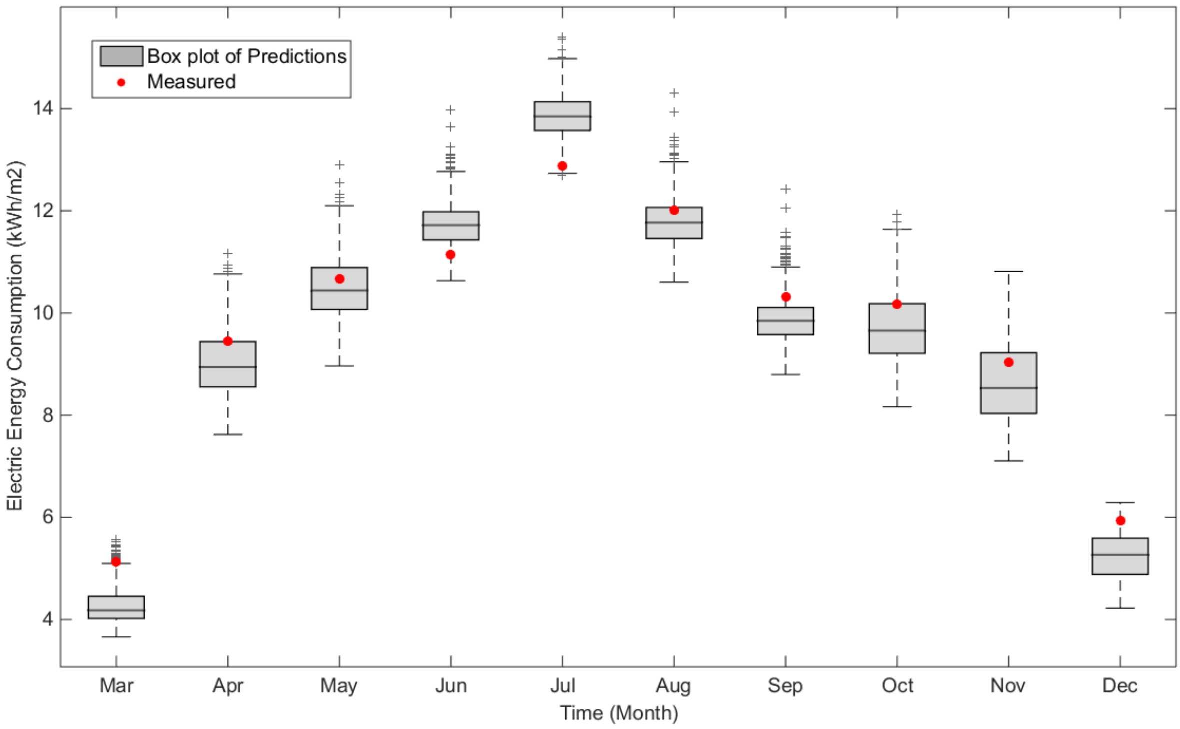

Figure 10 Comparison of model predictions with measured electricity consumption.

CONCLUSION

This paper has established a structured framework for modeling uncertainties in building performance simulation. Through a case study and using the application that was developed, 649 EnergyPlus input parameters were modeled as uncertain. In particular, a methodology for using observed data to quantify uncertainty was also demonstrated. It should be noted that Monte Carlo Analysis in building performance simulation is commonly carried out assuming a Gaussian, uniform or triangle distribution. This may not be representative of the uncertainties in actual conditions. Generating random samples from probability distributions that are based on observed data therefore provides the added benefit of being able to more accurately represent uncertainties. The underlying concept of this work is that by considering uncertainties and providing probabilistic predictions instead of the usual point estimates, decision makers would have greater confidence in simulation results.

At present, the database has been populated with parametric probability distributions that quantify uncertainties in material thermal properties, lighting power densities, equipment power densities, envelope infiltration and thermal properties of occupants and lights. However, uncertainties in HVAC and other building systems are still lacking. To increase the adoption of the proposed framework in building performance simulation, more research in the quantification of these uncertainties is required. More case studies is also needed to test the robustness of the proposed method and affirm the findings presented in this paper.

It is believed that with increasing ease in the application of Monte Carlo Analysis to building energy simulation programs such as EnergyPlus, its usage will become more widespread and commonplace.

REFERENCES

ASHRAE 2009. *ASHRAE handbook fundamentals*. American Society of Heating, Refrigerating and Air-Conditioning Engineers, Inc., Atlanta, GA, U.S.A.

Chantrasrisalai, C. and Fisher, D. E. 2007. Lighting heat gain parameters: Experimental results (rp-1282). *HVAC&R Research*, 13(2):305–324.

De Wit, S. 2004. Uncertainty in building simulation. In Malkawi, A. and Augenbroe, G., editors, *Advanced building simulation*. Taylor & Francis.

De Wit, S. and Augenbroe, G. 2002. Analysis of uncertainty in building design evaluations and its implications. *Energy and Buildings*, 34(9):951–958.

DOE 2014. Weather data for simulation. http://apps1.eere.energy.gov/buildings/energyplus/weatherdata_simulation.cfm.

Eisenhower, B., O'Neill, Z., Fonoberov, V. A., and Mezić, I. 2012. Uncertainty and sensitivity decomposition of building energy models. *Journal of Building Performance Simulation*, 5(3):171–184.

Emmerich, S. J. and Persily, A. K. 2005. Airtightness of commercial buildings in the us. In *26th AIVC Conference, Brussels, Belgium*.

EVO 2012. International performance measurement and verification protocol: Concepts and options for determining energy and water savings volume 1. *Efficiency Valuation Organization*.

Heo, Y., Choudhary, R., and Augenbroe, G. 2012. Calibration of building energy models for retrofit analysis under uncertainty. *Energy and Buildings*, 47:550–560.

Hopfe, C. J. and Hensen, J. L. 2011. Uncertainty analysis in building performance simulation for design support. *Energy and Buildings*, 43(10):2798–2805.

Lee, B. D., Sun, Y., Augenbroe, G., and Paredis, C. J. 2013. Towards better prediction of building performance: A workbench to analyze uncertainty in building simulation. In *Proceedings of the 13th IBPSA Building Simulation Conference*.

Macdonald, I. A. 2002. *Quantifying the effects of uncertainty in building simulation*. PhD thesis, University of Strathclyde.

Reddy, T. A. 2011. *Applied data analysis and modeling for energy engineers and scientists*. Springer Science & Business Media.

Sheppard, M. 2012. Fit all valid parametric probability distribution to data. *Matlab Central File Exchange*.

Struck, C. and Hensen, J. 2007. On supporting design decisions in conceptual design addressing specification uncertainties using performance simulation. In *Proceedings of the 10th IBPSA Building Simulation Conference*.

UIUC and LBNL 2014. Energyplus input output reference: the encyclopedic reference to energyplus input and output. *US Department of Energy*.

Wasserman, L. 2011. *All of statistics*. Springer Science & Business Media.