Why is the reliability of building simulation limited as a tool for evaluating energy conservation measures?

Nan Li ¹, Zheng Yang ², Burcin Becerik-Gerber ³, Chao Tang ⁴, Nanlin Chen ⁵

¹ Assistant Professor, Department of Construction Management, Tsinghua University, Beijing 100084, China (corresponding author, email: nanli@tsinghua.edu.cn)

² PhD Candidate, Sonny Astani Department of Civil and Environmental Engineering, University of Southern California, Los Angeles, CA 90089, U.S.A.

³ Associate Professor, Sonny Astani Department of Civil and Environmental Engineering, University of Southern California, Los Angeles, CA 90089, U.S.A.

^{4,5} Master Student, Sonny Astani Department of Civil and Environmental Engineering, University of Southern California, Los Angeles, CA 90089, U.S.A.

Abstract

Buildings account for approximately 32% of the total energy consumption globally and up to 40% in the developed countries, which makes buildings a prime target for energy conservation. Various energy conservation measures (ECMs) have been proposed to improve the energy efficiency in buildings, and these ECMs are usually designed and assessed using calibrated building energy models. However, there is empirical evidence that reveals noticeable discrepancies between simulated performances of ECMs reported in building energy models and their actual performances measured in buildings. This paper examines two possible causes of such discrepancies. Specifically, this paper tests the following two hypotheses: (1) using assumed occupancy data as opposed to actual occupancy data in building energy simulation reduces the reliability of estimated performance of demand-driven ECMs; and (2) using an energy model built for one ECM to cross estimate energy consumption of another ECM is statistically inaccurate. An educational building was used as a test bed. The results proved both hypotheses true, showing that estimations were more accurate and consistent for models calibrated using actual occupancy compared with those using assumed occupancy, and that cross-ECM estimation resulted in statistical inaccuracy. The findings indicated that current

building energy modeling methods have limited reliability in ECM performance assessment, and need to be improved to better support the design and implementation of ECMs in buildings.

Keywords: energy conservation measure (ECM); calibration; reliability; energy simulation; occupancy; cross estimation

1. Background

Buildings account for approximately 32% of the total energy consumption globally and up to 40% of the total energy consumption in the developed countries [1]. The fact that buildings are widely reported to be inefficient in terms of system operations [2,3] makes them a primary target for energy conservation. Considering the fact that existing buildings are generally in operation for 30 to 50 years and will account for 70% of the total building stock by 2050 [4], it is imperatively important to reduce energy consumption in existing buildings, for which various energy conservation measures (ECMs) have been proposed. For example, demand-response controls for lighting [5] and heating, ventilation, and air conditioning (HVAC) [6,7] are increasingly finding their way into building automation systems (BAS) and have reportedly achieved significant potential for energy savings. A more drastic approach for improving building energy efficiency is through energy retrofits, which usually involve the upgrade of building envelopes [8], building systems [9], and implementation of latest energy-efficient technologies [10].

Investigation of the feasibility and cost effectiveness of ECMs before their actual implementation is critical, especially that some of the ECMs might involve major retrofits or system changes and hence cannot be reversed once implemented, exerting long-term impacts on the energy performance of a building [11]. For ECMs that are tested or implemented by BAS and are hence reversible, their implementation may cause additional equipment depreciation and thermal discomfort to occupants, therefore, energy simulation could still be used as an inexpensive and non-intrusive alternative for ECM performance assessment. It has become the mainstream to employ building energy simulation models, believed to accurately reflect the physical characteristics and internal dynamics of the buildings, to quantify expected energy savings from the ECMs [12] and justify their applications [11]. First, ECMs are virtually

implemented by appropriately modifying building energy models, and the simulations are performed to examine the expected energy performance of buildings given these new changes. Repeated simulations can be performed to investigate potential opportunities for energy efficiencies [13], identify optimal strategies for building design [14] and daily building operations [15], and select among competing energy retrofit plans [16].

The reliability of reported performance of ECMs largely depends on the accuracy of the simulation model in representing the properties of the building it simulates [17], which in turn depends on how well the model is calibrated against available audit data [18]. The need for calibrating building energy simulation models in order to ensure reliable simulation results was first recognized in the 1970s [19], and since then it has been recognized as a fundamental factor in substantiating how well the models represent the characteristics of real buildings [20].

Calibration of building energy models is typically done by iteratively adjusting model parameters, based on the available audit data, until estimated energy consumption matches actual measurements [21-24] within tolerances dictated by certain criteria [25-27]. Uncertainty and sensitivity analyses are usually performed to identify the parameters that are the most impactful on key indicators predicted by the model of building energy performance, and these parameters are given higher weights in the calibration process. The audit data are usually composed of various sources, which are not necessarily coherent. The sources can be prioritized as follows in a descending order of reliability: sensor data, spot measurements, as-built drawings, design documents, and building standards [28]. The accuracy of calibrated models relies on various factors, such as the quantity of audit data and the complexity of building dynamics. More importantly, as a mostly trial-and-error process, the energy model calibration largely relies on the expert knowledge, past experience, statistical expertise, and engineering judgment [29]. The above calibration process is challenged by many researchers due to its ad-hoc and subjective nature, and recent efforts have focused on establishing formal and systematic calibration methodologies. As a result, a number of approaches, such as Bayesian calibration [12], reproducible calibration [10], evidence-based calibration [18,30], multi-stage calibration [31], and simultaneous multi-level calibration [32], have been proposed.

Despite the proposition of these new calibration approaches, there is accruing evidence showing significant discrepancies between estimated and actual energy savings [17,33,34]. Recent efforts

have focused on improving the reliability of building energy models for ECM assessments. For instance, Heo et al. generated normative models based on Bayesian calibration to account for the quantitative interventions of ECMs and provided probabilistic predictions for energy implications of different ECMs [11]; Parker et al. divided the calibration process into stages to comply with different ECMs [36]; Judkof et al. investigated the uncertainties and variability of energy consequences of different ECMs resulting from different simulation programs [37]; Reddy et al. used a group of models instead of a single model to estimate the effects of ECMs within range [29]; Coakley et al. proposed to update evidence in calibration to improve the representation of each parameter in order to increase the robustness of the model to different ECMs [28]. While these efforts have contributed to the reliability of energy simulation in ECM assessment, the improvements are generally limited and sometimes ECM-specific. Further efforts to reduce the discrepancy between estimated and actual energy savings of various ECMs are difficult, before the fundamental question of what have caused such discrepancy, in what way and to what extent, is clearly answered. Various factors may be responsible for the discrepancy, such as weather variations, faults in building control systems, occupant behaviors, occupancy schedules, and so on. This paper aims to investigate two possible causes of the discrepancy, including the use of actual occupancy data in energy modeling and the changes to building functionalities due to implementation of ECMs, and examine the overall reliability of building simulation as a tool for ECM assessment.

2. Research Motivation and Hypothesis

2.1 Challenges and motivations

The performance of ECMs in simulation lays the basis of their design and implementation, and a variety of building energy simulation approaches have been proposed in literature [38]. However, because of the discrepancies between actual buildings and their virtual representations, the optimality and expected energy savings of the ECMs as reported in simulations are not always met in practice. In fact, there is empirical evidence that reveals noticeable discrepancies between simulated and measured performances of ECMs [33,34]. From a theoretical point of view, there are two challenges that may prevent the simulated performances of the HVAC ECMs from being accurate.

First, occupancy is one of the most important influences on the thermal behavior of buildings [39]. It is a critical factor that determines the total and peak loads of HVAC systems and related energy consumption. Most building energy models built in prior research relied on assumed or simulated building occupancy, which inevitably deviates from the actual occupancy. Such simplification overlooks the impact of occupancy on building energy usage and, more importantly, cannot reflect the energy implications from interplays between occupancy patterns and the changes to building functionalities introduced by the ECMs. For instance, when estimating energy savings from an increase of indoor temperature set points in a warm climate, estimations from the simulation may be an underestimation, as occupants may react to the increased indoor temperatures by reducing the duration of their presence and hence their usage of the cooling service. Such implicit impacts of occupancy on building energy usage, not observable when assumed occupancy is used, highlight the need for using actual occupancy in assessing the effectiveness of HVAC ECMs. In fact, building occupancy detection itself is an area that has seen active research in the past decade. Different technologies have been proposed for occupancy detection, such as videos [40], CO₂ sensors [41], PIR sensors [42], Internet activity monitoring [43], or a combination of different technologies [44]. Despite the exploration of these solutions, large-scale occupancy detection has remained a challenging task, and has rarely been deployed at a building scale for evaluating ECMs.

Second, current modeling approaches calibrate models using the data collected when buildings are operated under a specific physical condition and operational strategy. The calibrated models are then used to estimate the energy consumption under various ECMs, assuming the buildings would have the same energy-consumption behaviors. However, buildings can be sensitive to physical conditions and operational strategies, and might behave differently under different ECMs. In current practice, energy models are generally calibrated using ad-hoc and subjective approaches [18] to simply approximate simulation results to energy ground truth. Conceptually, different ECMs could result in different simulation approximations, assumptions and simplifications [35]. Ideally, the model should be updated accordingly; In reality, however, such update is not possible in many cases because the estimation has to be done before ECMs are actually implemented, hence there is no ground truth data for model calibration. Consider a building as a mathematical operator that associates inputs (schedules, weather, etc.) with outputs (energy consumption). In current practice, once an operator is established (i.e., a model

is calibrated), the effectiveness of the ECM is evaluated by varying the inputs and estimating the associated outputs. However, the introduction of an ECM to the building may alter the functionality of the operator, which is assumed to be constant in current practice, and such alteration raises concerns about the significance of estimated outputs, in this case the effectiveness of ECMs. Although calibrated models can satisfy certain tolerance requirements, the uncertainty resulting from buildings being operated differently under various ECMs might cause challenges to the credibility of expected energy efficiencies [11]. Such uncertainty has not been given sufficient emphasis in the literature, and its impact has yet to be measured.

2.2 Research Hypotheses

To address the above-mentioned challenges, this paper proposes and validates the following two hypotheses.

Hypothesis one: Using assumed occupancy data as opposed to actual occupancy data in building energy simulation reduces the reliability of the estimated performance of an ECM.

For testing this hypothesis, two models (model 1 and model 2) for a test bed building were prepared and calibrated when the building was operated under an ECM (ECM 1). Under this ECM, the HVAC system in the building followed a fixed schedule. During the daytime (6:30 - 21:30 on workdays, and 7:00 - 21:30 on weekends), a constant temperature set point (22.8 °C) was maintained in all thermal zones by a proportional integral derivative (PID) controller, which dynamically adjusted the airflow damper and reheating valve of each zone. During the nighttime, the HVAC system was shut off except for providing minimum airflow to satisfy the ASHRAE compliance in order to save energy.

Model 1 was built based on the standard occupancy schedules in commercial prototype building models (medium office building type), developed by the Pacific Northwest National Laboratory and endorsed by the U.S. Department of Energy. These schedules are embedded in various simulation software tools including EnergyPlus. Prepared based on the ASHRAE 90.1 2013, commercial prototype building models represent 80% of the commercial building floor area in the United States and are widely used by researchers and professionals in a variety of ways for building energy simulation purposes [45,46]. Model 2 was built based on the actual occupancy schedules observed in the building. For collecting the actual occupancy data, an occupancy

detection system proposed by the authors was used [47]. High-resolution occupancy data of the test bed building was collected as part of the audit data and used in the modeling process. Both models, after developed and calibrated with the same procedure and the same audit data (except for the occupancy data), are used to estimate the energy consumption of the building after ECM 1 is introduced. Their respective estimation accuracies are compared to validate this hypothesis.

Hypothesis two: Using an energy model built for one ECM to cross estimate energy consumption of another ECM is statistically inaccurate.

This hypothesis is motivated by the second challenge discussed in Section 2.1. For testing this hypothesis, model 3 of the test bed building was prepared and calibrated. Unlike the model 2 that was built based on operational data collected when the test bed building was operated under ECM 1, model 3 was built based on the operational data collected after another ECM (ECM 2) was implemented. Under ECM 2, actual zone-level occupancy was used to drive the HVAC operations. During the daytime, for occupied zones, a constant temperature set point (22.8 °C) was maintained by the PID controller. If a zone stayed vacant for a minimum of 15 minutes, the temperature set point was increased to 25.5 °C until the room became occupied again. The HVAC system was shut off during the nighttime, and no heating, cooling or ventilation services are provided except for minimum airflow to satisfy the ASHRAE compliance. After both models 2 and 3 were developed and calibrated with the same procedure, actual occupancy data and non-operational data (e.g., drawings and specifications), they were used to cross estimate the energy consumption of the building under each other's ECM. The reported estimation accuracies were analyzed to test hypothesis two.

3. Test Bed Building

The test bed building modeled in this paper is the Ralph & Goldy Lewis Hall (RGL), a typical building on the University of Southern California (USC) campus near downtown Los Angeles, California. It is a three-story building with a footprint of 3,735 m². It has 89 mechanically ventilated rooms including offices, classrooms, conference rooms, auditoriums and lobbies. A total of 64 wired temperature sensors and 50 wireless sensor units are installed to monitor the indoor environment of the building. Each wireless sensor unit has a stand-alone single-board microcontroller with integrated support for wireless communications, and is comprised of the

following sensors: a light sensor, a sound sensor, a motion sensor, a CO₂ sensor, a temperature sensor, a relative humidity sensor, a PIR sensor, and a door switch sensor. Data are automatically queried every one minute, time stamped, and stored in an SQL database. The energy consumption by various building systems such as HVAC, lighting, receptacle and mechanical is metered and recorded in a building energy management system (BEMS).

The 89 mechanically conditioned rooms are divided into 67 mechanical zones. The indoor climate of 64 zones is serviced by variable air volume (VAV) units, and fan coil units service the rest of the three zones. A VAV box is responsible for regulating the ventilation in the thermal zone with conditioned air, and reheating the air with hot water supplied by the boilers if the zone needs heating instead of cooling. The conditioned air is supplied to the VAVs by air handler units (AHUs) using fans and ductwork. There are two AHUs in the building, each servicing one side of the building with similar sizes of service areas. The AHUs take in outside air, mix it with returned air from the building, and cool down the mixed air to 12.8 °C with chilled water supplied by the chillers. The HVAC energy consumption in the building can be decomposed to heating and cooling energy consumption, used by chillers and boilers to generate chilled and hot water, respectively, and ventilation energy consumption, used by AHUs and their embedded fans to distribute conditioned air in the building.

4. Methodology

4.1 Occupancy Modeling

A real-time occupancy modeling approach introduced by the authors in [47] was deployed in the test bed building. The fundamental principle is that occupancy regularly influences and interferes with the ambient environment, therefore, by mathematically or statistically bridging occupancy ground truth and ambient factor variations through supervised learning using historical data, future ambient data can be analyzed to output corresponding occupancy outcomes. A total of 11 ambient sensor variables, based on the sensor data reported by the aforementioned wireless sensor units, were used in occupancy modeling. These variables could be categorized into three types: (1) instant variables that show the instant output of a sensor at the time the data are queried, including light level, binary motion, CO₂ concentration, temperature, humidity, binary PIR, and door status (open/close); (2) count variables that sum number of times a sensor's output changes in the last minute, (motion count net, PIR count net,

and door count net); (3) average variables that show average value of a sensor's output over a certain period of time (sound average - every 5 seconds). Several machine-learning algorithms were tested. It was found that pruned decision tree had the best performance in solving the classification problem for occupancy modeling, demonstrating the relationship that each ambient factor was responsible for classifying part of the instances and occupancy classes were determined by the sequential consideration of ambient factors. After the occupancies of all individual rooms were estimated, the occupancy of a specific thermal zone was determined by integrating the occupancy of the associated rooms. Only if all the rooms within one zone were vacant, the zone was considered as vacant. By following the above methodology, occupancies of 50 rooms were calculated and recorded (accuracy of 97%) for a period of four months from January 1 to April 30, 2013 with three-minute granularity. Rooms not covered by the sensor units, including a number of offices, an auditorium and three classrooms, used assumed occupancy.

4.2 Building Model Calibration

The aforementioned three models were prepared and calibrated following the procedure described below. As illustrated in Figures 1 and 2, first, the geometry of the test bed building was modeled in SketchUp [48] representing the physical building structure. The geometry was then imported to OpenStudio [49], in which most model parameters were set based on the available evidence, such as the information collected from site visits and operation and maintenance (O&M) manuals. The model was then saved as an IDF file, and imported to EnergyPlus [50] for setting advanced parameters and simulating the energy performance of the building in a given period. Energy consumption ground truth data, such as electricity and gas consumed by the entire HVAC system, were collected and calculated for four months between January 1 and April 30, 2013 using the BEMS of the test bed building. The ECM 2 was implemented on the east side of the second and third floors and the ECM 1 in the rest of the building between February 22 and March 31, 2013. The ECM 1 was implemented in the entire building during the rest of the four months. The collected actual occupancy data were converted to EnergyPlus model inputs with the schedule editing function in OpenStudio. The assumed occupancy schedules, written in fractional schedule format [45,46], were overwritten with actual occupancy data, using 0 and 1 to indicate vacancy and occupancy, respectively. The

resulting actual occupancy schedules were then embedded in IDF files generated in OpenStudio, and used in simulations in EnergyPlus.

Figure 1: Modeling geometry and basic parameter setting in SketchUp

Figure 2: Setting detailed parameters in OpenStudio

The guideline used for setting the types and values of model parameters is summarized as follows. The parameters related to the construction layers and materials were defined based on as-built architectural, structural and mechanical drawings, and site surveys. The parameters related to thermal loads associated with lighting and electrical equipment were defined based on recommendations in OpenStudio. There were 560 parameters, whose values could not be determined based on available audit data. The values of these uncertain parameters were set based on the modeler's best guesses and tuned later in the calibration process. The outside weather profile was linked to typical meteorological year (TMY) data [51] for Los Angeles. Assumed occupancy schedules based on the standard occupancy schedules in commercial prototype building models (medium office building type) were used in model 1, and actual occupancy schedules based on the occupancy modeling approach explained in section 4.1 were used in models 2 and 3.

Building energy model calibration processes are defined in several standards and guidelines [25-27] and are followed in this paper. First, a sensitivity analysis was carried out [52] to identify the top influential parameters among the 560 uncertain parameters on thermal loads of the building. For performing the sensitivity analysis, each parameter was assigned a value domain, based on its definition in EnergyPlus input/output reference guide [53], and a probability distribution of its value, based on the following guidelines: (1) A normal distribution was assigned to parameters that have preferred values but no specific value domains in the reference guide, such as "central cooling design supply air humidity ratio"; (2) A uniform distribution was assigned to the parameters that have specific value domains but no preferred values in the reference guide, such as "cooling minimum air flow"; (3) A triangle distribution was assigned to parameters that have both preferred values and specific domains in the reference guide, such as "fan efficiency"; and (4) A discrete distribution was assigned to non-numerical parameters, such as "boiler flow mode". A total of 8000 models were then generated, by

randomly sampling the above parameters within their respective domains using the commonly adopted Latin Hypercube Sample (LHS) technique.

All models were then simulated in EnergyPlus using parallel computing provided by the USC high-performance computing (HPC) facility. Since the parameters were sampled independently without considering the relationships between them, the integrity of some models was violated, leading to early termination of the simulation or incomplete and erroneous simulation results. Around 50% of the models completed the simulation and yielded valid HVAC energy consumption data. The parameter values of these models, and their corresponding HVAC energy consumption data were then analyzed using the analysis of variance (ANOVA) method, which provided an assessment of the influence of each parameter on HVAC energy consumption.

Based on the ANOVA analysis results, all uncertain parameters were ranked in a descending order of their influence. The top 5% or 28 influential parameters were identified and further tuned in the calibration process. A genetic algorithm (GA)-based process was used in tuning these top influential parameters, due to the global and random search ability of GA as well as its fast convergence. More importantly, the GA-based tuning process does not require rich modeling experience or thorough knowledge of all of the thermodynamics and heat balance algorithms in the simulation program, and is not subject to influence by the characteristics of input parameters.

In the tuning process, a model that represents one combination of the values of all top influential parameters was considered an individual, and a set of combinations was considered a population and a subset of the entire search space. Fitness is defined as the measure to evaluate the performance of an individual, and is comprised of two commonly used tolerances including MBE (mean bias error) and CV (RMSE) (coefficient of variation of the root mean square error). In this paper, the values of daily MBE and CV (RMSE) were equally weighted (50% for each) to calculate the fitness of the individual through the process of energy simulation. Specifically, the genes with accurate and consistent estimations were selected into the next generation and got further tuned. Three genetic operators were defined to implement the GA algorithm. The first operator was selection, which selected individuals from current population for later breeding. A roulette wheel selection rule was used, where the probability for an individual to get selected was proportional to its fitness value. The second operator was crossover, which at a 100%

crossover rate interchanged the value of a randomly selected parameter between two individuals. The last genetic operator is mutation, which updates the input parameters based on its range and probability density function. The mutation rate was set to p [0.01, 0.1] and the parameter(s) to be mutated was selected. The locations for both crossover and mutation were determined by the uniform growth probability density function so that parameters with higher influence were more likely to be selected. The population was then updated and the fitness was re-calculated to determine whether any satisfactory individuals existed in the population. If not, the genetic operations were iterated until a satisfactory model was found. The process of implementing GA to tune the influential parameters is shown in Figure 3.

Figure 3: GA-based parameter tuning process

An individual was considered satisfactory when its fitness value was within preset tolerances. Table 1 summarizes the tolerances of MBE and CV (RMSE) specified in relevant standards [23-25]. However, the energy consumption estimation was carried out on a daily basis in this paper. Daily estimation is the highest granularity allowed by the energy consumption data available in the BEMS. It is also an appropriate granularity that allows a fine-grained measurement of energy consumption not achievable by annual or monthly estimation, while avoiding drastic fluctuations of hourly estimation due to changes of ambient factors, such as outside temperature. There is no commonly used tolerance for daily estimation. Hence, based on existing monthly and hourly tolerances, a daily tolerance of 8% for MBE and 20% for CV (RMSE) is introduced in this paper.

Table 1: Existing tolerances for building energy model calibration

An important factor to be considered in the calibration process is the amount of training data. Building energy simulation is an excessively complicated process and involves relatively too many parameters compared to the number of training data points used for model calibration. Therefore, the models may be overcalibrated by overfitting its performance to the training data, which would result in low estimation accuracy. To avoid the overfitting issue, an analysis of the trends of MBE and CV (RMSE) with different days of training data was conducted for all three models. It was found that the minimum days of training data were 21, 27 and 24 for models 1, 2

and 3, respectively. Therefore, all models were calibrated using 27 days of training data. The settings of the three models are summarized in Table 2.

Table 2: Settings of the three models

5. Assessment

Models 1 and 3 were used to estimate the HVAC energy consumption in the test bed building between March 21 and March 31 when the ECM 2 was implemented. This process mimicked what practitioners would do in practice, i.e., using an energy model calibrated with assumed occupancy to assess the energy consumption of the building operated under another ECM. Model 1 represented models used in practice (using assumed occupancy, calibrated under ECM 1 to estimate ECM 2), while model 3 represented models that should be used in ideal cases (using actual occupancy, calibrated under ECM 2 to estimate ECM 2). The estimation results are shown in Figure 4. Compared to model 3, model 1 yielded lower accuracies with much higher fluctuations. More importantly, model 1 tended to overestimate the energy consumption while model 3 tended to underestimate it. The comparison confirmed that the current approach to ECM assessment may lead to inaccurate and misleading results. The potential reasons for such discrepancy between the two models are examined in this section.

Figure 4: Daily estimation accuracies of models 1 and 3

5.1 Similarity Assessment

The performance of the calibrated models is inherently determined by the uncertain parameters that are tuned during the calibration process. The introduction of actual occupancy and the implementation of ECMs are likely to impact the influence of these parameters and hence impact the characteristics and performance of the calibrated models.

The ANOVA analysis explained in section 4.2 ranked the parameters based on their influences observed in a sensitivity analysis. For either hypothesis, the characteristic similarity of the two models could be measured by the similarity of their parameter rankings, which was calculated based on a lexical order induced by the parameter ranking of model 2. The results showed that, when all 560 parameters were taken into consideration, the characteristic similarities between

models 1 and 2, and between models 2 and 3, were 32.8% and 39.4%, respectively. This finding suggests that the models had significantly different parameter rankings and hence essentially distinct characteristics. Moreover, when the union of top 5% influential parameters of all three models, which included 63 parameters, was taken into consideration, the characteristic similarities further reduced to 20.1% and 25.4%, suggesting that the impacts were more significant among top influential parameters.

The uncertain parameters could be grouped into five clusters, including construction details, HVAC, equipment, lighting, occupancy, and simulation control, based on parameter definitions in EnergyPlus input/output reference guide [52]. The influence of a cluster is defined as the average influence of all parameters in the cluster. The normalized cluster influences were compared between the models to further assess their characteristic similarity. As can be seen in Figures 5 and 6, the relative influence of the clusters were highly consistent between models 1 and 2, and between models 2 and 3. The results suggest that, although highly distinct in terms of the influence ranking of individual parameters, the models still shared certain fundamental characteristics that were not subject to the impact of the introduction of actual occupancy and the implementation of ECMs.

Figure 5: Comparison of cluster influences between models 1 and 2

Figure 6: Comparison of cluster influences between models 2 and 3

5.2 Assessment of Hypothesis One

Both models 1 and 2 were used to estimate the daily energy consumption during March 21 - 31, 2013, when the building was operated under ECM 2. First, the assumed occupancy profile and actual occupancy profile were compared. Their similarity was examined on a daily basis, by calculating the average difference between assumed occupancy data points, which were probabilistic values between 0 and 1, and actual occupancy data points, which were deterministic 0/1 binary values, throughout the day, for both the calibration period (27 days) and estimation period (11 days). The results are shown in Figure 7. Weekends are highlighted with squares in the figure. The average daily similarity of the two occupancy profiles was 78.39%, with a standard deviation of 9.16%. The similarity was higher and more consistent on weekends,

with an average of 88.42% and a standard deviation of 2.87%, compared to weekdays, with an average of 73.77% and a standard deviation of 7.07%.

Figure 7: Daily similarity of assumed occupancy and actual occupancy

An examination of the daily estimation accuracies of models 1 and 2 was then carried out. The results showed that the MBE values were all positive for model 1 and all negative for model 2, which suggested that model 1 consistently overestimated the energy consumption, while model 2 consistently underestimated the energy consumption. One possible reason for such discrepancy is that model 1 used assumed occupancy that likely overestimated the usage of the room and hence overestimated the heat loads in a cooling-dominant climate zone (e.g., Los Angeles).

The daily estimation accuracies are shown in Figure 8. As can be seen in the figure, estimations by model 1 had noticeably lower accuracies and higher fluctuations. A paired sample t-test between the two groups of MBE values reported a t value of 10.597 and a p value of 0.000, which rejected the null hypothesis at a significance level of 95% and suggested that the estimation accuracy of model 2 is statistically higher than that of model 1. The estimations by model 1 had a CV (RMSE) value of 17.9%, which was 3.9% higher than the one by model 2, suggesting that estimation accuracy of model 2 was statistically more consistent than that of model 1.

Figure 8: Daily estimation accuracies of models 1 and 2

5.3 Assessment of Hypothesis Two

First, both models 2 and 3 were used to estimate the daily energy consumption between March 21 and March 31, 2013, when the ECM 2 was implemented. Both models tended to underestimate the energy consumption. The daily estimations reported by the two models are illustrated in Figure 9. As shown in the figure, the daily estimation accuracies by model 2, with an MBE value of -8.0% and a CV (RMSE) value of 17.9%, were consistently lower than those by model 3, with a MBE value of -6.6% and a CV (RMSE) value of 14.4%. Such difference was further proven by a paired sample t-test of the daily estimation accuracies. The test reported a t value

of 9.915 and a p value of 0.000, which rejected the null hypothesis at a significance level of 95% and suggested that estimations by model 3 were statistically more accurate than those of model 2.

Figure 9: Daily estimation accuracies of models 2 and 3 March 21 and March 31

The same assessment was carried out between April 1 and April 30, when the ECM 1 was implemented in the test bed building. Both models again tended to underestimate the energy consumption. The daily estimations (absolute values) reported by the two models are illustrated in Figure 10. As shown in the figure, the daily estimation accuracies by model 2, with a MBE value of -6.8% and a CV (RMSE) value of 14.7%, were consistently higher than those by model 3, with a MBE value of -8.1% and a CV (RMSE) value of 18.7%. Such difference was further proven by a paired sample t-test. The test reports a t value of 4.918 and a p value of 0.000, which rejected the null hypothesis at a significance level of 95% and suggested that estimations by model 2 were statistically more accurate than those of model 3.

Figure 10: Daily estimation accuracies of models 2 and 3 between April 1 and April 30

The above results indicate that the energy consumption estimations are statistically more accurate when the building is operated under the same ECM during both modeling training and estimation periods. In other words, the results support hypothesis two of this paper, that using an energy model built for one ECM to cross estimate energy consumption of another ECM is statistically inaccurate.

6. Discussion and Conclusions

Building energy simulation is widely used in ECM assessment. However, there is accruing evidence showing significant discrepancies between estimated energy savings reported by simulation models and actual energy savings achieved in practice, and the reasons for such discrepancy are yet to be answered. This paper examines two potential causes of the discrepancy by testing two hypotheses, including: (1) using assumed occupancy data as opposed to actual occupancy data in building energy simulation reduces the reliability of estimated performances of ECMs; and (2) using an energy model built for one ECM to cross estimate

energy consumption of another ECM is statistically inaccurate. Three energy models were built and calibrated for a test bed building using different types of occupancy and when the building was operated under different ECMs.

The results reported in section 5 show that the model, calibrated using actual occupancy, can provide statistically more accurate estimation of building energy consumption for a new ECM than the model calibrated using assumed occupancy. Therefore, the first hypothesis of this paper is proven to be true, suggesting that building energy simulation can be made a more reliable tool for assessing the effectiveness of ECMs by basing the calibration of energy models on actual occupancy data. The impact of using actual occupancy in the performance of calibrated models likely lies in the fact that occupancy is closely related to the internal thermal loads caused by occupants' metabolic heat gains and their use of lighting and appliances. These thermal loads in turn cause responses of the HVAC system, as it works to maintain a given temperature set point. Building energy models that are calibrated using actual occupancy are adaptive to patterns of frequent occupancy changes and associated thermal loads in the built environment, as well as patterns of responses of an HVAC system. Therefore, these models would yield better estimations of energy consumption and hence better assessment of ECMs.

The test results also show that cross-ECM estimation of building energy consumption is statistically inaccurate, which proves true the second hypothesis of this paper. A potential reason for such inaccuracy is that the implementation of the ECMs causes essential changes to the way that internal thermal loads are regulated. The HVAC system might behave differently, which may not be predictable by models that are trained with data that represent old behavioral patterns. This finding suggests that models built for current ECMs may be unreliable tools for assessing new ECMs that are planned to be implemented. Despite the prevailing use of cross-ECM estimation for ECM assessment in current practice, it needs to be emphasized that assessments performed with current approaches could be inaccurate and misleading, and should be interpreted with caution.

The results illustrate that the introduction of actual occupancy and implementation of new ECMs can change the characteristics of the models and cause statistical variations of the estimation accuracies. The reliability of using building energy simulation in ECM assessment is therefore challenged. These results concur with the concerns raised in prior research [32,54].

Although building energy simulation is the most accurate and widely used among various ways to evaluate ECMs [55], it still bears unreliability and inconsistency, since calibrated models may only conform to the ground truth instead of being accurate representation of buildings [18]. The differences in simulation approximations, assumptions, and simplifications cannot be eliminated through calibration [35]. In particular, when the ECMs are related to the controls of building systems, like the ones implemented in this paper, the values of certain interactive input parameters may not be accurately modeled leading to biased simulation [55,56], thus the models are not generalizable and representative enough to adapt to different operating conditions and load profiles resulting from the implementations of different ECMs [57,58].

The findings are further supported by the low characteristic similarity of the models in both hypotheses. It indicates that the models are inherently distinct, and that one model cannot be used as a perfect substitute for the other model as practitioners have been doing in practice. It also needs to be noted that the relative influences of parameter clusters are generally consistent between the models, suggesting that certain fundamental characteristics of the models are determined by factors other than occupancy or HVAC control strategies. These consistent characteristics explain why the discrepancies of models' performances are far less drastic than the discrepancy of the influence ranking of individual parameters.

The assessment of the effectiveness of ECMs is needed in a variety of scenarios, such as application for building permits, planning of building retrofits, and documentation for LEED (leadership in energy and environmental design) building certification. The fact that both hypotheses in this paper are proven true highlights that the reliability of current assessment is questionable. This paper examines two factors in one test bed building. Admittedly, the findings are not conclusive, because there may be other factors that can further explain the discrepancy between simulated and actual ECM performance, and it is unknown whether the hypotheses are always true when they are tested in different types of buildings, at different locations, and in different seasons of the year. Still, current findings suggest that there is a need for rethinking of the current approaches that have limited reliability in ECM performance assessment, and that current assessments of ECMs should be interpreted and used with caution until better assessment approaches become available. In addition, there are two observations that could provide further suggestions to current ECM assessment practice. First, the occupancy data used for calibrating model 1 and that for model 2 were partially different, as nearly half of the rooms

were not covered by sensor units and they shared the same assumed occupancy data across the two models. Therefore, the observed impact could have been larger if actual occupancy data were available for and applied to all indoor spaces in the test bed building, which further emphasizes that actual occupancy data are of significant importance in using building energy models for ECM assessment. Second, it was observed that the model calibrated using actual occupancy tended to underestimate the energy consumption of the test building. In practice, such underestimation may lead to overoptimistic projection of potential savings from implementing an ECM and hence requires prudence of building owners or managers when they make investment plans.

While the reliability issue of using building energy models for assessing ECMs has been well noticed in the literature, there is no easy solution. Two major causes of the unreliability are revealed and examined in this paper, and extensive future research is needed to design better assessment approaches that account for the impact of actual occupancy and implementation of ECMs. Since actual occupancy data are not accessible in most buildings, research can be done to develop cost-efficient, non-intrusive occupancy sensing tools so that large-scale indoor occupancy monitoring, which has numerous other benefits besides energy modeling, becomes feasible. Alternatively, data mining algorithms can be developed in future research to estimate the occupancy using accessible data, such as operational data of HVAC systems, whose status is continuously updated in reaction to, among other factors, changing occupancy and associated thermal loads. It is also challenging to eliminate the impact of implementation of ECMs illustrated in the second hypothesis. Further efforts are needed to develop approaches to analyzing potential impact of physical and functional changes of building systems on energy consumption, to predicting behavioral changes of occupants in response to changes in indoor climate that could be brought by the ECMs, and to integrating these factors in the developing and calibration process of building energy models.

Acknowledgments

This material is based upon work supported by the National Science Foundation under Grant No. 1351701. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National

Science Foundation. The occupancy data and building performance data were collected under DE-EE0004019. Computation for the work described in this paper was partly supported by the University of Southern California's Center for High-Performance Computing (hpc.usc.edu).

References

[1] DOE. Building Energy Data Book. <http://buildingsdatabook.eren.doe.gov/>. Last accessed 2014, Dec 10.

[2] MA Piette, SK Kinney, P Haves. Analysis of an information monitoring and diagnostic system to improve building operations, *Energy Build.* 33 (2001) 783-91.

[3] MM Ardehali, TF Smith, JM House, CJ Klaassen. Building energy use and control problems: An assessment of case studies, *ASHRAE TRANSACTIONS TECHNICAL AND SYMPOSIUM PAPERS*, (2003).

[4] M Kelly. Energy efficiency, resilience to future climates and long-term sustainability: The role of the built environment, *Phil Trans R Soc A.* 368 (2010) 1083-9-1083-9.

[5] SA Raziei, H Mohscnian-Had. Optimal demand response capacity of automatic lighting control, 2013 IEEE PES Innovative Smart Grid Technologies Conference, ISGT 2013, February 24, 2013 - February 27, (2013), 2013.

[6] VL Erickson, AE Cerpa. Occupancy based demand response HVAC control strategy, 2nd ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Buildings, BuildSys'10, November 2, 2010 - November 2, (2010), 2010.

[7] S Goyal, HA Ingley, P Barooah. Occupancy-based zone-climate control for energy-efficient buildings: Complexity vs. performance, *Appl. Energy.* 106 (2013) 209-221.

[8] B Gucyeter, HM Gunaydin. Optimization of an envelope retrofit strategy for an existing office building, *Energy Build.* 55 (2012) 647-59.

[9] S Cho, JS Haberl. Development of a high-performance office building simulation model for a hot and humid climate, 11th International IBPSA Conference - Building Simulation 2009, BS 2009, July 27, 2007 - July 30, (2009), 2007.

[10] R Dadioti, S Rees. Operational performance and calibrated model analysis of a large low-energy university building, 13th Conference of International Building Performance Simulation Association, BS 2013, August 26, 2013 - August 28, (2013), 2013.

[11] Y Heo, R Choudhary, GA Augenbroe. Calibration of building energy models for retrofit analysis under uncertainty, *eEnergy Build.* 47 (2012) 550-60.

586 [12] Z O'Neill, B Eisenhower, S Yuan, T Bailey, S Narayanan, V Fonoberov. Modeling and
587 calibration of energy models for a DoD building, 2011 ASHRAE Annual Conference, June 25, 2011
588 - June 29, (2011), 2011.

589 [13] E Azar, CC Menassa. A comprehensive analysis of the impact of occupancy parameters in
590 energy simulation of office buildings, *Energy Build.* 55 (2012) 841-853.

591 [14] A Nguyen, S Reiter, P Rigo. A review on simulation-based optimization methods applied to
592 building performance analysis, *Appl.Energy.* 113 (2014) 1043-1058.

593 [15] Z Yang, N Li, B Becerik-Gerber, M Orosz. A Systematic Approach to Occupancy Modeling in
594 Ambient Sensor Rich Office Environments, *Simulation.* 90 (2014) 960-977.

595 [16] Y Pan, Z Huang, G Wu. Calibrated building energy simulation and its application in a high-
596 rise commercial building in Shanghai, *Energy Build.* 39 (2007) 651-657.

597 [17] D Coakley, P Raftery, M Keane. A review of methods to match building energy simulation
598 models to measured data, *Renewable and Sustainable Energy Reviews.* 37 (2014) 123-141.

599 [18] P Raftery, M Keane, J O'Donnell. Calibrating whole building energy models: An evidence-
600 based methodology, *Energy Build.* 43 (2011) 2356-64.

601 [19] JM Ayres, E Stamper. Historical development of building energy calculations, *ASHRAE J.* 37
602 (1995) 8pp-8pp.

603 [20] G Liu, M Liu. A rapid calibration procedure and case study for simplified simulation models
604 of commonly used HVAC systems, *Build.Environ.* 46 (2011) 409-420.

605 [21] D Zhou, SH Park. Simulation-assisted management and control over building energy
606 efficiency - A case study, 2011 2nd International Conference on Advances in Energy Engineering,
607 ICAEE 2011, December 27, 2011 - December 28, (2012), 2011.

608 [22] S Goyal, HA Ingley, P Barooah. Zone-level control algorithms based on occupancy
609 information for energy efficient buildings, 2012 American Control Conference - ACC 2012, (2012),
610 27-29 June 2012.

611 [23] N Fumo, P Mago, R Luck. Methodology to estimate building energy consumption using
612 EnergyPlus Benchmark Models, *Energy Build.* 42 (2010) 2331-7.

613 [24] X Zhou, Y Zhu, C Xia, H Chen. Simulation-based method to assess building energy
614 consumption level at operation stage, *Building Simulation 2007, BS 2007*, September 3, 2007 -
615 September 6, (2007), 2007.

616 [25] ASHRAE, ASHRAE Guideline 14-2002: Measurement of Energy and Demand Savings, *ASHRAE*,
617 Atlanta, GA, 2002,.

618 [26] Efficiency Valuation Organisation, International Performance Measurement and Verification
619 Protocol, 2007,.

620 [27] US Department Of Energy, M&V Guidelines: Measurement and Verification for Federal
621 Energy Projects Version 3.0, 2008,.

622 [28] D Coakley, P Raftery, P Molloy, G White. Calibration of a detailed Bes model to measured
623 data using an evidence-based analytical optimisation approach, 12th Conference of
624 International Building Performance Simulation Association Building Simulation 2011, BS 2011,
625 November 14, 2011 - November 16, (2011), 2011.

626 [29] TA Reddy, I Maor, C Panjapornpon. Calibrating detailed building energy simulation
627 programs with measured data - Part I: General methodology (RP-1051), HVAC and R Research.
628 13 (2007) 221-241.

629 [30] P Raftery, M Keane, A Costa. Calibrating whole building energy models: detailed case study
630 using hourly measured data, Energy Build. 43 (2011) 3666-79.

631 [31] G Mustafaraj, D Marini, A Costa, M Keane. Model calibration for building energy efficiency
632 simulation, Appl.Energy. 130 (2014) 72-85.

633 [32] Z Yang, B Becerik-Gerber. A model calibration framework for simultaneous multi-level
634 building energy simulation, Appl.Energy. 149 (2015) 415-431.

635 [33] VA Smith, R Young, AJ Spata, D Fisher. Virtual vs. real: modeling the energy performance of
636 a quick service restaurant, ASHRAE Trans. 105 (1999) 1122-1137.

637 [34] Tamara Ann Kuiken. Energy Savings in a Commercial Building with Daylighting Controls:
638 Empirical Study and DOE-2 Validation, Iowa State University, (2002).

639 [35] T Maile, V Bazjanac, M Fischer. A method to compare simulated and measured data to
640 assess building energy performance, Build.Environ. 56 (2012) 241-251.

641 [36] J Parker, P Cropper, L Sha. A calibrated whole building simulation approach to assessing
642 retrofit options for Birmingham airport, First Building Simulation and Optimization Conference,
643 Loughborough, UK, Sep 10-11, 2012.

644 [37] R Judkoff, D Wortman, B O'Doherty, J Burch. A methodology for validating building energy
645 analysis simulations, NREL/TP-550-42059, National Renewable Energy Laboratory, (2008). [38] N
646 Fumo. A review on the basics of building energy estimation, Renewable and Sustainable Energy
647 Reviews. 31 (2014) 53-60. [39] F Oldewurtel, D Sturzenegger, M Morari. Importance of
648 occupancy information for building climate control, Appl.Energy. 101 (2013) 521-532.

649 [40] H Wang, Q Jia, C Song, R Yuan, X Guan. Estimation of occupancy level in indoor environment
650 based on heterogeneous information fusion, 2010 49th IEEE Conference on Decision and
651 Control, CDC 2010, December 15, 2010 - December 17, (2010), 2010.

652 [41] Z Sun, S Wang, Z Ma. In-situ implementation and validation of a CO₂-based adaptive
653 demand-controlled ventilation strategy in a multi-zone office building, Build.Environ. 46 (2011)
654 124-133.

655 [42] Y Agarwal, B Balaji, R Gupta, J Lyles, M Wei, T Weng. Occupancy-driven energy management
656 for smart building automation, 2nd ACM Workshop on Embedded Sensing Systems for Energy-
657 Efficiency in Buildings, BuildSys'10, November 2, 2010 - November 2, (2010), 2010.

658 [43] R Melfi, B Rosenblum, B Nordman, K Christensen. Measuring building occupancy using
659 existing network infrastructure, 2011 International Green Computing Conference, IGCC 2011,
660 July 25, 2011 - July 28, (2011), 2011.

661 [44] S Meyn, A Surana, Y Lin, SM Oggianu, S Narayanan, TA Frewen. A sensor-utility-network
662 method for estimation of occupancy in buildings, 48th IEEE Conference on Decision and Control
663 held jointly with 2009 28th Chinese Control Conference, CDC/CCC 2009, December 15, 2009 -
664 December 18, (2009), 2009.

665 [45] S Goel M, Rosenberg, R Athalye, Y Xie, W Wang, R Hart, J Zhang, V Mendon. Enhancements
666 to ASHRAE standard 90.1 prototype building models. PNNL-23269, Pacific Northwest National
667 Laboratory, (2014).

668 [46] M Halverson, R Athalye, M Rosenberg, Y Xie, W Wang, R Hart, J Zhang, S Goel, V Mendon.
669 ANSI/ASHRAE/IES Standard 90.1-2013 Determination of Energy Savings: Quantitative Analysis,
670 PNNL-23479, Pacific Northwest National Laboratory, (2014).

671 [47] Y Zheng, N Li, B Becerik-Gerber. A Non-Intrusive Occupancy Monitoring System for Demand
672 Driven HVAC Operations, Proceedings of The Construction Research Congress 2012, (2012), May
673 21-23, West Lafayette, IN.

674 [48] Trimble Buildings. SketchUp. www.sketchup.com. Last accessed 2014, Dec 23.

675 [49] NREL. OpenStudio. <https://www.openstudio.net>. Last accessed 2014, Dec 23.

676 [50] U.S. Department of Energy. EnergyPlus. <http://apps1.eere.energy.gov/buildings/energyplus>.
677 Last accessed 2014, Dec 23.

678 [51] National Renewable Energy Laboratory. National Solar Radiation Data Base.
679 http://rredc.nrel.gov/solar/old_data/nsrdb/1991-2005/tmy3/. Last accessed 2015, Jan 20.

680 [52] B Eisenhower, Z O'Neill, VA Fonoferov, I Mezic. Uncertainty and sensitivity decomposition
681 of building energy models, Journal of Building Performance Simulation. 5 (2012) 171-184.

682 [53] US Department of Energy. The Encyclopedic Reference to EnergyPlus Input and Output.
683 <http://apps1.eere.energy.gov/buildings/energyplus/pdfs/inputoutputreference.pdf>. Last
684 accessed 2015, Jan 20.

- 685 [54] Y Zhang, Z O'Neill, B Dong, G Augenbroe. Comparisons of Inverse Modeling Approaches for
686 Predicting Building Energy Performance, Build.Environ. 86 (2015) 177-190.
- 687 [55] TA Reddy. Literature review on calibration of building energy simulation programs: uses,
688 problems, procedures, uncertainty and tools, ASHRAE Transactions, 112, 1 (2006) 226.
- 689 [56] TA Reddy, I Maor, C Panjapornpon. Calibrating detailed building energy simulation
690 programs with measured data - Part II: Application to three case study office buildings (RP-1051),
691 HVAC and R Research. 13 (2007) 243-265.
- 692 [57] L Pérez-Lombard, J Ortiz, JF Coronel, IR Maestre. A review of HVAC systems requirements in
693 building energy regulations, Energy Build. 43 (2011) 255-268.
- 694 [58] M Manfren, N Aste, R Moshksar. Calibration and uncertainty analysis for computer models -
695 A meta-model based approach for integrated building energy simulation, Appl.Energy. 103
696 (2013) 627-641.

Figure List

Figure 1: Modeling geometry and basic parameter setting in SketchUp

Figure 2: Setting detailed parameters in OpenStudio

Figure 3: GA-based parameter tuning process

Figure 4: Daily estimation accuracies of models 1 and 3

Figure 5: Comparison of cluster influences between models 1 and 2

Figure 6: Comparison of cluster influences between models 2 and 3

Figure 7: Daily similarity of assumed occupancy and actual occupancy

Figure 8: Daily estimation accuracies of models 1 and 2

Figure 9: Daily estimation accuracies of models 2 and 3 March 21 and March 31

Figure 10: Daily estimation accuracies of models 2 and 3 between April 1 and April 30

Figure 1: Modeling geometry and basic parameter setting in SketchUp

Figure 2: Setting detailed parameters in OpenStudio

Figure 3: GA-based parameter tuning process

Figure 4: Daily estimation accuracies of models 1 and 3

Figure 5: Comparison of cluster influences between models 1 and 2

Figure 6: Comparison of cluster influences between models 2 and 3

Figure 7: Daily similarity of assumed occupancy and actual occupancy

Figure 8: Daily estimation accuracies of models 1 and 2

Figure 9: Daily estimation accuracies of models 2 and 3 March 21 and March 31

Figure 10: Daily estimation accuracies of models 2 and 3 between April 1 and April 30

Table List

Table 1: Existing tolerances for building energy model calibration

Table 2: Settings of the three models

Table 1: Existing tolerances for building energy model calibration

Standards	Monthly (%)		Hourly (%)	
	MBE	CV (RMSE)	MBE	CV (RMSE)
ASHRAE	5	15	10	30
IPMVP	20	-	5	20
FEMP	5	15	10	30

Table 2: Settings of the three models

Model #	Type of occupancy	Calibration period	ECM in calibration period	Estimation period to mimic current practice	Estimation period to test H1	Estimation period to test H2 under ECM 1	Estimation period to test H2 under ECM 2
1	Assumed occupancy	Jan 15 – Feb 10, 2013 (27 days)	ECM 1	Mar 21 – 31, 2013 (11 days)	Mar 21 – 31, 2013 (11 days)	--	--
2	Actual occupancy	Jan 15 – Feb 10, 2013 (27 days)	ECM 1	--		April 1 – 30, 2013 (30 days)	March 21 – 31, 2013 (11 days)
3	Actual occupancy	Feb 22 – March 20, 2013 (27 days)	ECM 2	Mar 21 – 31, 2013 (11 days)	--		