Evaluation of “Autotune” Calibration against Manual Calibration of Building Energy Models

Gaurav Chaudhary^{#1}, Joshua New^{*2Θ}, Jibonananda Sanyal^{*3}, Piljae Im^{*4}, Zheng O’Neill^{@5}, Vishal Garg^{†6}

[#]*Department of Architecture and Planning,*

Indian Institute of Technology Roorkee, India

¹*gcapsuap@iitr.ac.in*

^{*}*Oak Ridge National Laboratory, Oak Ridge, TN, USA*

²*newjr@ornl.gov*

³*sanyalj@ornl.gov*

⁴*imp1@ornl.gov*

[@]*Department of Mechanical Engineering,*

The University of Alabama Tuscaloosa, AL, USA

⁵*zoneill@eng.ua.edu*

[†]*Centre for IT in Building Science,*

International Institute of Information Technology Hyderabad, India

⁶*vishal@iiit.ac.in*

Abstract

This paper demonstrates the application of Autotune, a methodology aimed at automatically producing calibrated building energy models using measured data, in two case studies. In the first case, a building model is de-tuned by deliberately injecting faults into more than 60 parameters. This model was then calibrated using Autotune and its accuracy with respect to the original model was evaluated in terms of the industry-standard normalized mean bias error and coefficient of variation of root mean squared error metrics set forth in ASHRAE Guideline 14. In addition to whole-building energy consumption, outputs including lighting, plug load profiles, HVAC energy consumption, zone temperatures, and other variables were analyzed. In the second case, Autotune calibration is compared directly to experts’ manual calibration of an emulated-occupancy, full-size residential building with comparable calibration results in much less time. The paper concludes with a discussion of the key strengths and weaknesses of auto-calibration approaches.

Θ **Corresponding Author:** Postal address: P.O. Box 2008, MS-6324 Oak Ridge, TN 37831, USA; e-mail address: newjr@ornl.gov ; tel.: (865) 241-8783; fax: (865) 241-4152.

Keywords

- Autotune
- Building energy modeling
- Calibration
- Energy efficient buildings
- Automated Calibration

Acronyms		
ASHRAE	American Society of Heating, Refrigerating, and Air-Conditioning Engineers	
BEM	Building Energy Models	
BEPS	Building Energy Performance Simulation	
COP	Coefficient of Performance	
ECMs	Energy Conservation Measures	
ESCO	Energy Service Company	
HVAC	Heating, ventilation, and air conditioning	
IPMVP	International Performance Measurement and Verification Protocol	
M&V	Measurement and Verification	
NSGA II	Nondominated Sorting Genetic Algorithm - II	
SHGC	Solar Heat Gain Coefficient	
WWR	Window-to-Wall Ratio	
Nomenclature		
CV(RMSE)	Coefficient of Variation of Root Mean Square Error	[%]
MBE	Mean Bias Error	[%]

1. Introduction

India is experiencing an unprecedented construction boom. It doubled its floor space between 2001 and 2005 and is expected to add 35 billion m² of new buildings by 2050 [1]. As of 2013, the building sector in India is consuming nearly 40% of the electricity used and is expected to use 76% by 2040 [2]. In the United States (US) and the United Kingdom (UK), 39% of the total energy is consumed by buildings, and that figure is projected to increase annually by 1.7% to 2025 [3]. It is estimated that, by 2030, 63% of building floor space in China and around 65% in India will be in urban areas. It is projected that buildings will soon be the single largest energy-consuming sector in the world. Therefore, improving building energy efficiency can be one of the fastest, easiest, and most cost-effective paths toward reducing energy use worldwide [3]. In order to realize energy efficient buildings, measurement and simulation are established as two of the most common approaches [4,5].

Energy Service Company (ESCO) industry revenue in the United States was about USD \$5.3 billion in 2011 with energy efficiency projects accounting for about 85% of that revenue [6]. It is expected to double in size from ~\$6 billion (2013) to \$11-\$15 billion (2020) [7]. ESCOs undertake energy retrofits of existing buildings through energy performance contracts that typically guarantee savings as part of their service. Retrofitting existing buildings require calibrated baseline energy models in order to design and optimize retrofits by defining specific energy conservation measures (ECMs) that yield the best return on investment. However, due to the man-hours and cost required to create a calibrated model, it is employed only in large projects.

A generalized, fast, flexible, and easy-to-use automated building energy model calibration tool could, with sufficient accuracy and market confidence, facilitate market penetration of building modeling into smaller buildings by significantly reducing the costs to develop calibrated models and enhance the cost-effectiveness of software-optimized retrofit projects. It would also significantly improve the state of the art in energy savings measurement and verification (M&V) for performance contracting of ESCOs, and could potentially optimize control strategies in real-time [8]. If automated calibration technologies are able to increase the retrofit market by even 1% while expanding cost-effective reach of retrofitting into smaller buildings, it would amount to a cumulative energy savings of 27.4 TBtus/year (9.16 GW) [9].

In addition, using a calibrated model of buildings is well aligned with next generation Demand Side Management (DSM) technologies. This could facilitate building owners and facility management staff to make informed decisions about their energy usage, in turn reducing emissions of greenhouse gases. This ability to control electricity usage could translate into as much as USD \$59 billion in social benefit by 2019 [10].

This paper presents two case studies for comparisons between a manual calibration and an automated calibration methodology which has yet to enter the market in its full form. This unique study provides in-depth analysis and all necessary data to allow a comparison of individual calibration methods, along with business-relevant details that could facilitate increased adoption of calibration technologies by the ESCO industry and utility companies.

2. Building Energy Modeling Overview

Building energy modeling is an approach for calculating building energy consumption. EnergyPlus is an open-source whole-building energy simulation program that engineers, architects, and researchers use to model energy use in buildings. Calibration of a building energy model to utility data is often essential for the simulation engine to produce realistic estimates of energy use. These calibrated energy models are useful for commissioning of building systems, estimations of potential savings from applying energy conservation measures, and measurement and verification of building retrofit projects. A typical building energy model could have an average of around 3,000 parameters that could be calibrated. Manual calibration for only dozens of parameters could require months. Autotune is a method developed from the application of multi-objective algorithms to millions of simulations. It mines hundreds of terabytes of data on high performance computing resources to quantitatively identify and optimize the best technique for calibrating building energy models using measured data (such as utility bills, sub-meter data, and/or sensor data). Such methodology

creates models accurately and robustly by deriving near-optimal input parameter values in a systematic, automated, and repeatable fashion.

Building simulation is the use of software to predict energy use of a building and is reliant upon a virtual representation (i.e. building model) and models of interaction for energy-related variables (i.e. energy simulation). It is widely used for integrating heat and mass transfer, environmental data, and load-HVAC (heating, ventilation, and air conditioning) interactions. It can be used to periodically update energy performance estimates of various building systems (e.g., HVAC) [11-13]. Z. Yang [14] outlines several advantages of simulation over field experiments: (1) allows evaluation of system performance when field experiments are infeasible; (2) estimates energy savings of various ECMs before being implemented; (3) is less time consuming and less expensive; (4) can be performed repeatedly to check effect of a parameter; (5) permits control of factors that cannot be controlled in a field experiment (e.g., weather conditions); (6) is non-intrusive to building occupants; (7) output of performance indicators that would be difficult to measure; and (8) easy interpretations of results.

To reduce the amount of energy consumed and slow the growth of energy demand in buildings, it is important to understand the energy distribution throughout a building [15]. The accurate modeling of a building can serve as an important tool for investigating and understanding this energy distribution. Building Energy Performance Simulation models (BEPS), also known as Building Energy Models (BEM), may be used to compare the cost-effectiveness of ECMs—changes that can be made to a building to save energy, either in the design phase of a new building or in retrofitting existing buildings. However, because of the complexity of the built environment and the large number of parameters interacting with one another, it can quickly become difficult to accurately model a real building [16].

According to a model taxonomy by Saltelli [17], BEPS models can fall on a spectrum from diagnostic (used to understand a law) to prognostic (used to predict systemic behavior from a supposedly understood law) as well as on a spectrum from data-driven (statistical properties from observations) to law-driven (typically a physics equation). BEPS models typically involve properties governing system-relevant variables (e.g., mass balance, energy balance, heat transfer, conductivity, etc.) and can be divided into three categories.

- Black-box models (data-driven models) are mathematical (e.g., statistical) models constructed from training data. Better models can be realized through a plethora of high-quality data, but typically lack physical meaning in their mapping of influential input parameters (e.g., occupancy or weather) to measured outputs. The relative development time of these models is short, although they can be brittle and require re-training to accommodate small changes in a building. Examples include models created by traditional regression [18-20], artificial neural networks [21-23], or support vector machines [24-26].
- Grey-box models often use parameters identified from physical systems, can use decision tree [27,28] or Fourier series techniques [29-31], and can account for changes caused by simplified input parameters.
- White-box models (including law-driven models such as BEPS) [8,31,32] employ fully descriptive algorithms to explicitly link the physical building, system, and environmental parameters. This can provide the most detailed building performance characteristics and predictive strength is determined by the quality of input data as well as the algorithmic accuracy to the relevant physics. There are a variety of BEPS tools including DOE-2, ESPr, EnergyPlus and TRNSYS [34-37]; specifications such as Annex 53, IPMVP and ASHRAE Guideline 14 [38-40]; and manuals for use of these models in building design and operation [41]. It has been established that BEPS have significant applied energy benefits in the area of: 1) energy consumption prediction, 2) energy savings for return-on-investment (ROI) optimization, and 3) establishing baseline building performance [42].

There are more than 120 BEPS software tools and 20 popular ones have been evaluated for their various strengths and weaknesses toward realistically modeling the whole-building physics related to energy use [43]. Among these, the U.S. Department of Energy has invested over \$65 million since 1995 in the development of EnergyPlus, its flagship whole-building energy simulation tool. It maintains some functionality of its predecessor DOE2 regarding prediction of energy use given hourly weather information, building geometry, material properties, and HVAC description [44]. EnergyPlus has a large support community of active users and is used internationally to model HVAC, lighting, and water consumption. It employs a heat balance

method for building heating/cooling load calculation, where governing state-space equations are used for building envelope heat transfer. [45,46]. A major challenge for these whole-building energy simulation tools is how to enable realistic modeling of a building while remaining cost-effective to use when EnergyPlus typically accommodates more than 3,000 input parameters for the average building.

Although even inaccurate models can be tremendously useful, specific business applications such as retrofit projects, commissioning, and operations management require models that are sufficiently accurate according to established guidelines [47]. However, we can achieve more accurate and reliable results from a building energy model by changing model input to reconcile model output with measured data via a process known as “calibration” [15]. Modification of simulation algorithms to more closely match real-world physics can also help make models more accurate, but the process of tuning virtual building models to match real data remains the easiest, most straightforward way to meet calibration guidelines. Unfortunately, the building energy model calibration field has not advanced significantly, as this tuning process remains an “art” [48] that less than 5% of practitioners report enjoying.

At present, BEPS models are under-utilized by the market for the primary reason that they lack the credibility [49]. We attempt to demonstrate that the lack of education, quality, and reproducibility underlying this credibility issue can largely be addressed by improvements in input quality via the calibration. Automated calibration can assist with several related BEPS issues: 1) Standards – lack of understanding and consistent use of standardized methods, 2) Expense – time, learning curve, and cost required to synthesize accurate inputs for a building, and 3) Integration – poor integration among 3D modeling (e.g., Autodesk Revit, ArchiCAD) and BEPS simulation packages (e.g., EnergyPlus, TRNSYS and Modelica) [50].

Despite the advantages of BEPS, expected energy savings from ECMs reported in simulations typically do not match those measured in actual buildings due to the discrepancies between actual buildings and their virtual representations. Numerous studies [51-56] have revealed noticeable differences frequently ranging from 23 to 97% on a monthly basis. Driving factors for inputs that contribute to simulation uncertainty and discrepancies in building energy usage include: 1) the variability of occupant behavior [57-59], 2) operation and maintenance changes, 3) climate impacts [60,61], 4) alterations in the indoor environmental conditions, 5) internal heat gains, and 6) building equipment and system configuration and efficiency [38]. Driving factors for simulation uncertainties include simplifications of building systems, assumptions underlying thermal processes, and algorithmic differences used in simulation programs [19,62]. Predicted energy savings and actual energy savings can vary by as much as 11 times between different simulation models for high performance buildings incorporating passive design; this understandably leads to users, often architects or engineers, to question the validity of simulation results beyond a relative trend [63]. The building model must accurately represent the actual behavior of the building if BEPS models are to be used with any degree of confidence [50]. We propose that such a level of precision can be achieved through the model calibration, and result in a useful model with minimal efforts that closely matches measured building performance, even for energy-related variables not used in the calibration procedure.

The remainder of this paper is structured as follows. In Section 3, a brief literature study of current and past calibration approaches, and need of the “Autotune” calibration approach are presented. It is followed by Section 4 which contains details about the Autotune methodology and how it works. In Section 5, we provide the step-by-step methodology for using Autotune calibration to recover a manually-calibrated model from a detuned model. And in Section 6, the performance of Autotune is presented in a competition with a manual calibration of a residential building by an expert modeler and calibrator [64].

3. Building Energy Model Calibration

Model calibration is, mathematically, a search for a highly-parameterized model in an under-determined search space. This parameterization includes a large number of independent and interdependent input parameters representing the correlations and dynamic interactions among building envelope thermal conditions, HVAC responses, interior impacts (e.g., light related internal heat gains), and exterior impacts (e.g., solar radiation) which typically cannot all be obtained empirically [19]. Collecting necessary data and determining the precise simulation input values for a special BEPS use case requires significant time, effort, and cost [65]. Three reasons to conduct model calibration include: 1) improving accuracy of building energy

performance models [66], 2) provide insights into a building's thermal or electric hourly load shapes [67], and 3) better predict energy savings of an energy conservation measure [68].

Typical targets of building energy model calibration include the ability to match the simulated energy consumption [58,69,70], indoor air temperature [53,71], operation conditions of HVAC equipment [68], or heating/cooling loads [53,68,97] to measured meter/sensor data of the building, on a specific time scale. This time scale usually depends on the purpose of calibration and the resolution of available sensor data. Calibration on an hourly or sub-hourly time scale is more difficult and time-consuming than an annual or monthly time scale.

During the early days of BEPS, simple percentage differences between measured data and simulated data were the primary metrics of the comparison [72-74]. However, this method allows a compensation effect whereby overestimations (i.e., positive errors) are canceled out by underestimations (negative errors). Therefore, a method was proposed [72] which adopted standard statistical indices that better represent the performance of a building energy model [75-77].

- Mean bias error (MBE) (%): MBE is a non-dimensional bias measure (i.e., a sum of errors) between measured and simulated data for each timestep (typically monthly or hourly). The MBE is a good indicator of the overall (positive or negative) bias in the model. It captures the mean difference between measured and simulated data points. However, positive bias compensates for negative bias (the cancellation effect) over the entire time period (typically one year). Hence, a further measure of model error is required:

$$MBE(\%) = \frac{\sum_{i=1}^{N_p} (m_i - s_i)}{\sum_{i=1}^{N_p} (m_i)} \quad (1)$$

- Root mean square error (RMSE) (%): The RMSE is a measure of the variability of the data. For every timestep, the difference in each pair of data points is calculated and squared. The sum of squared errors (SSE) is then added for the total period and divided by the respective number of timesteps, yielding the mean squared error (MSE). A square root of the result is then reported as the RMSE.
- Coefficient of variation of RMSE CV(RMSE) (%): This index enables determination of how well a model fits the data by capturing offsetting errors between measured and simulated data. It does not suffer from the cancellation effect except within each timestep [46].

$$CV(RMSE)(\%) = \frac{\sqrt{\frac{\sum_{i=1}^{N_p} (m_i - s_i)^2}{N_p}}}{\bar{m}}, \quad (2)$$

where m_i and s_i are the measured and simulated data points, respectively, for each model instance ' i '; N_p is the number of data points at interval ' p ' (i.e., $N_{monthly}=12$, $N_{hourly}=8760$); and $\bar{m}$ is the average of the measured data points.

Currently, in the U.S, most of practitioners and researcher follow three guidelines listed in Table 1 for the validation of BEPS models. For example, BEPS models are generally considered calibrated if they meet the criteria defined by ASHRAE Guideline 14.

Table 1. Two different statistical indices have been used to evaluate if a BEPS model is sufficiently calibrated: the normalized mean bias error (NMBE) defined by Equation (1), and the coefficient of variance of the root-mean-square error (CV-RMSE) defined by Equation (2). Acceptance ranges for these indices for monthly and hourly data is defined by different standards/guidelines.

	Monthly criteria (%)		Hourly criteria (%)	
Standard/guideline	MBE	CV(RMSE)	MBE	CV(RMSE)
ASHRAE Guideline 14 [40]	5	15	10	30
International Performance Measurement and Verification Protocol [39]	20	–	10	20

Federal Energy Management Program [78]	5	15	10	30
--	---	----	----	----

Since calibration is a mathematically under-constrained problem, a model that meets these conditions will not be unique. That is, many, models of the same building with significantly different inputs can be considered calibrated. Note also that calibration criteria relate only to simulation output matching measured data; they do not account for uncertainty or directly for inaccurate input parameters [50].

An extensive body of work exists on BEPS calibration with many important works documented in the literature review of calibration programs, tools and techniques by Reddy [79], Fumo [80] and D. Coakley et al. [50]. These reviews reveal that there is no generally adopted methodology by which building energy models should be calibrated [50,79,80] due to different purpose-driven requirements, configurations of building systems, availability of data, and levels of calibration experience. Common calibration techniques [70,79] include calibration based on: (i) manual, iterative and pragmatic intervention, (ii) a suite of informative, graphically-comparative displays, (iii) special tests using analytical test procedures, and (iv) mathematical methods for calibration. Instances of these approaches fall along a spectrum of completely manual to fully automated techniques [50].

Manual calibration approaches are iterative and pragmatic in nature and hence involves tuning or refining initial input parameters in a heuristic manner, relying heavily on the experience and expertise of the modeler [81,75]. Manual calibration utilizes building characteristics data from audits, energy usage data, and zone condition monitoring to gain detailed knowledge of the physical and operational characteristics of the building [69,52,82].

Graphical techniques have been widely used in manual calibration to visually show the differences between measured and computed results, following manual parameter tuning [75,83-86]. A process has been developed to use graphical signatures of heating and cooling load to calibrate a building energy model [87-90]. This was enhanced to demonstrate a rapid two-stage calibration procedure for simplified BEPS models based on use of calibration signatures [91].

A number of studies have incorporated a systematic and evidence-based model development at the core of the calibration process [15,75,92-96]. A bottom-up calibration approach using hourly, sub-metered electric use and HVAC cooling/heating load has been demonstrated [42]. Also, a calibration study of a public office building combining a building energy audit, model sensitivity analysis, and manual tuning of influential parameters has been conducted [97].

Manual calibration requires skill; moreover, it is a time consuming and costly process to run most simulation engines, wait for the new outputs, evaluate the match to measured data, and then change more inputs. Inadequate maintenance, sensor calibration, and/or quality assurance may cause poor data quality which can lead to suboptimal model calibration. However, manual calibration combines human intelligence, expertise, and experience into a trial-and-error process which has the potential to lead to a calibrated model that is more reliable and closer to the actual building [86,98,99]. Because the calibrated models depend upon subjective modeler expertise and limited time, the credibility of the process can always be questioned.

Unlike manual calibration, automated calibration commonly relies on mathematical and statistical methods of calibration. Most of these techniques utilize an optimization function to reduce the difference between measured and simulated data (e.g., modify inputs to iteratively minimize mean square error between measure and simulated data). This can be augmented with a penalty function to reduce the likelihood of deviating too far from the base-case [100-102]. Many systematic calibration methods also include parameter estimations as well as determinations of uncertainty associated with input parameters in the final model [101].

Bayesian calibration methods [103,104] can be used to incorporate uncertainties in the calibration process. These uncertainties may be propagated through the model using probabilistic sensitivity analysis [105]. Existing automated calibration processes have included Bayesian approach [106], meta-models [19], sensitivity analysis with meta-models [107], multi-level parameters and simulation levels with discrepancy analysis [13], and identification of influential parameters with sensitivity analysis prior to simulated annealing

[108]. More recently, works include a data driven probabilistic graphic model to predict building energy performance [109], a lightweight Bayesian calibration of dynamic models that accounts for model parameter uncertainties [110], whether monthly energy output data can be treated as informative or uninformative data in Bayesian calibration [111], and the NSGA-II genetic algorithm [112].

Automated calibration can be treated as a multi-objective optimization problem, which is a mathematically-based rather than physically-based approach (i.e., it will be able to closely match simulation results to measured data, but may not necessarily match the actual building's physical properties.) Current automated calibration usually must run a simulation for each of the, potentially thousands, generated building models to determine how well each matches measured data; this can require large amounts of computation (e.g. cloud computing or supercomputing) to complete the calibration process within an acceptable time frame. While most automated calibration systems either stop once the error rate drops below a specific target value, or at a specific time limit, there is a trade-off between the number of variables and the time required to reach a specific error rate. With additional calibration parameters, there is a combinatorial increase in complexity to find the "best" model due to confounding variables that could impact energy use similarly. A decrease in error rate during the calibration period is often similar in form to that shown in Figure 7. However, the calibration performance for a given amount of time is highly variable, and a specific error rate may be impossible for a given set of calibration parameters and ranges. A generalized understanding and characterization of the number, types, and ranges of calibration parameters on calibration time and error rate remains an open research problem.

Issues with BEPS models calibration are highlighted in a Rocky Mountain Institute (RMI) study [49], 1) Lack of Standards: Current guidelines specify only broad ranges of allowable error for building energy models, but not for issues such as input uncertainty or zone-level inaccuracy. 2) Expense: There is difficulty in accommodating a lack of standard tool-chains, interoperable file formats, and obtaining high-resolution energy use data. 3) Simplification: There are a lack of methods for developing robust models which are neither over-specified (i.e., too many inputs) nor under-determined (i.e., too few validation points) for the calibration. 4) Inputs: There is a lack of high-quality input data required for detailed models. 5) Uncertainty: lack of confidence can result from a simulation's single answer or an inability to effectively utilize uncertainty ranges. 6) Identification: There are challenges with identifying the underlying causes of discrepancies between model predictions and measured data.

In the current literature, there is no single calibration methodology which is seen to be generally adopted for the calibration of BEPS models [80,50]. The lack of any formal methodology can result in findings that are "highly dependent on the personal judgment of the analyst doing the calibration" [79]. There is an explicit lack of tool-chains which can be used to calibrate a BEPS model to guideline criteria. Current calibration techniques are both time consuming and costly, where manual calibration requires the time of an experienced building energy analyst while automated calibration requires computing power and time to complete the calibration process. To address these shortcomings, this paper presents a novel methodology for conducting automated model calibration based on evolutionary computation algorithms, called "Autotune". This approach was refined from a systematic testing of multiple calibration algorithms to require less computing power, time, and human intervention to robustly calibrate different types of buildings. The goal is to achieve lower error rates than guideline requirements and enable model predictions to match the actual building. In addition, the Autotune performance is comparable to a human expert. The Autotune project aims to save the time building modelers spend tweaking building input parameters to match ground-truth data by providing an Autotune easy button for a standard desktop computer which intelligently adjusts model inputs.

4. Autotune

The Autotune methodology [113] was developed at Oak Ridge National Laboratory (ORNL) to calibrate a BEPS model to measured data in a completely automated fashion. Supercomputing systems were used to run millions of parametric simulations for different building types across various climate zones for training machine learning agents [114]. The Autotune project leveraged many supercomputers, including the one that was then the world's fastest (the 299,000-core Jaguar/Titan), to generate the simulation data for the project. The project also made use of the latest advances in web-oriented database storage for queryable and publicly-sharable storage of 156 inputs and 96 outputs at 15-minute resolution for 8 million EnergyPlus simulations. In

an effort to promote open science, the Autotune project has made a small portion of the 200+ TB (26.9 trillion data points) of EnergyPlus simulation data publicly available at bit.ly/1nVevp5 [48].

Building systems, equipment, and materials are continually becoming more complicated and diverse with more monitoring and control points. Likewise, BEPS model algorithms are evolving to more thoroughly model existing systems and capture new equipment technologies. Therefore, there is a need to mitigate this increasing complexity by relying on cost-effective, intelligent algorithms to calibrate BEPS models while using as much data as is available. [115]

Figure 1 illustrates Autotune's workflow for replacing the inaccurate and expensive manual calibration process with software intelligence informed by machine learning to fully-automate calibration for the EnergyPlus simulation engine.

Fig. 1. The Autotune workflow leverages an EnergyPlus model of a building and measured data for calibration. When initial building models are compared with utility bills, they never match. Thus an expert must calibrate the model in an expensive feedback loop. Autotune provides an easy, fully automated method to calibrate the model using rigorously tested multi-objective optimization algorithms.

The Autotune project [113] aims to mitigate increasing complexity with an automated process and has previously demonstrated calibration results for envelope parameters using monthly utility data [48]. The Autotune project aims to replace art with science and expensive human time with cheap computing time. Autotune identified and optimized evolutionary computation to calibrate model inputs using any sources of measured data that can map to simulation engine output.

Reduced-order models (ROMs), typically generated using regression analysis applied to full simulation engines, can allow for energy use estimates while requiring fewer input parameters [116,117]. Such models have lower statistical degrees of freedom that could be calibrated in less time or more accurately. However, such models are generally not robust to the wider array of input variables of interest to most practitioners [116]. For ESCOs that replace lighting, reduced order models or spreadsheet estimates may be sufficient; however, ESCOs involving replacement of HVAC equipment would ideally be interested in calibrating any parameter that could affect heating and cooling loads within the building(s). Autotune, instead, uses a full simulation model and allows users the freedom to select a subset of input parameters for calibration. In this way, for example, a 12-parameter model could be used to articulate a full simulation model. To mitigate the complexity of selecting from ~3,000 parameters, Autotune has identified 47-470 parameters most important for each building type using sensitivity analysis that a user can load to quickly get started and down-select further based on the practitioner's use case (bit.ly/29J3HHM). Autotune also supports "grouping" individual parameters, which acts as a generator layer in which a single degree of freedom can change the value of dozens of input parameters.

4.1. Working of Autotune

A typical BEPS model in EnergyPlus has approximately 3,000 input parameters that must be specified for a given building. The search space in such a problem is extremely large. The search space for a 3,000-parameter building would contain 2^{3000} possibilities, even if each parameter were a simple binary value (e.g., yes or no). This is much larger than the estimated 10^{80} number of subatomic particles in the observable universe. However, the actual size of the search space is effectively infinite, because many of the parameters are continuous-valued [115]. A common approach to such search problems, successfully applied in previous building model tuning efforts, [48] is evolutionary computation (EC). Evolutionary algorithms have been shown to efficiently search such extremely large spaces. They generally avoid the problem of local optima by maintaining a population of possible solutions, rather than performing a strictly gradient-based approach [118].

EC [119-122] is a stochastic search algorithm that attempts to mimic biological evolution by maintaining a set of candidate solutions referred to as a “population.” In this case, a candidate solution is a building with a set of building model parameters to be tuned. Each candidate solution is evaluated to determine its fitness, a problem-dependent measure of how well it solves the problem [123]. To evaluate a candidate solution, its corresponding EnergyPlus input data file (*.idf) model is constructed and passed to the EnergyPlus simulation engine. These output measures from the simulation results are then compared with actual measured data from the building. The resulting measure of accuracy is used as the fitness value for the candidate solution. During each cycle of the evolutionary computation or generation, individuals from the population are selected for modification and are modified using evolutionary operators (recombination and/or mutation) to produce new solutions. Then some set of existing solutions is allowed to continue to the next generation [122]. The particular evolutionary operators used in this work were heuristic crossover, Gaussian mutation, tournament selection, and generational replacement. In Figure 2, two buildings (i.e. parent solutions) are recombined to produce two offspring by using genetic operators of crossover and mutation; principles of fitness and natural selection are used to drive subsequent offspring toward a close match between simulation output and measured data.

A previous study [115] introduces Autotune and shares 5 of the 19 experiments used to refine the calibration methodology as it was being developed; this was done by reducing combinatorial complexity to a computationally tractable search problem that could scale to handle hundreds of parameters and dozens-to-millions of measured data points. Particularly, the study explores methods for speeding up the calibration through parallelization and simulating only portions of a year in the context of overall calibration accuracy for two models (primitive and refined) generated for a single building during a manual calibration.

In contrast, this study introduces two unique test cases that: (1) demonstrate practical use of automated calibration, (2) allow detailed, quantitative analysis of calibration performance, and (3) quantify business-relevant gains in terms of accuracy and man-hours saved by such technologies. In Case Study 1 (Section 5), a practical step-by-step guide shows how to use Autotune with in-depth analysis of a calibration algorithm’s ability to recover deliberate faults manually inserted into 63 parameters. In Case Study 2 (Section 6), a detailed EnergyPlus model of a real-world, emulated-occupancy residential building is calibrated using hourly electric meter data. This new calibration using Autotune is then compared with a detailed manual calibration of the same model which was done previously by Im et al. [64].

Fig. 2. Evolution progresses as candidate solutions undergo the effects of evolutionary operators, such as mutation. In this illustration, the two parent solutions are recombined to produce two offspring as candidate solutions for the next generation. The offspring shown here are produced by averaging each component and then adding Gaussian noise.

4.2. Different Platforms through which Autotune Can Be Used

The Autotune code is freely available as open source code on the GitHub repository of ORNL's Building Technologies Research and Integration Center (bit.ly/autotune_code). There are detailed instructions made available in *README.md* for the various ways one may run Autotune, but we summarize a few here:

- Windows operating system: A directory named "demo" contains all files necessary to run Autotune from a real-world, 303-variable calibration study of a building with sub-hourly (15-minute) data. The files include a base case building model (myidf.idf and optional myschedule.csv), parameters to calibrate (myparams.csv), weather data (myweather.epw), and measured data (myuserdata.csv). After proper installation, as documented, a user should be able to download and invoke an Autotune calibration by running autotune.bat.
- Large demo on a virtual machine (any operating system): This larger demo illustrates the full functionality of Autotune as a stand-alone, browser-based application (website) or as an interactive web service. A Linux (Ubuntu) image, loadable via virtual machine, is provided for quickly running the full functionality without having to download and configure all required software packages.
- Server installation (Windows or Linux): This method provides the same functionality as the virtual machine but in a native operating system environment.

Autotune was developed to be simulation engine agnostic, and could be modified to apply to any buildings simulation engine (e.g. TRNSYS, DOE-2) or any software for which inputs need to be tuned such that software output matches measured data.

5. Case Study 1: Comparative Methodology using a De-tuned Model

In this section, we provide the step-by-step methodology for using Autotune calibration. A building model was created, and deliberate faults were injected to de-tune it. This faulty model was then calibrated to evaluate Autotune’s performance in recovering the original model.

5.1 About the Model

The model is for an 800 m² (8,611 ft²), rectangular, one-story office building divided into five conditioned zones (four perimeter, one core) as shown in Figure 3. Windows are on all four facades for a window-to-wall ratio of 40%, and there are eight skylights in the core zone, each measuring 1.5 m². All windows are shaded by 3 m wide overhangs. Wall, roof, and floor construction matches the constructions defined in ASHRAE’s 30% savings advanced energy design guide provided by ASHRAE Special Project 102 Committee. The infiltration rate is set at 0.1 air changes per hour. The air conditioning in the five zones is provided by a fan-powered variable-air-volume system with an outside air economizer, hot water reheat coils, and central chilled water cooling coil. The central plant includes a single hot water boiler, an electric compression chiller with a water-cooled condenser. A summary of the building characteristics is provided in Table 2.

Table 2. The office building used in this study is divided into five conditioned zones. All the constructions are according to ASHRAE’s 30% savings advanced energy design guide.

Characteristics	Description
Floor area	800 m ² (8,611 ft ²)
Construction	Wall, roof, and floor construction matches the constructions defined in ASHRAE’s 30% savings advanced energy design guide
Windows	Single pane, 40% WWR for windows on all four facades: U = 2.716 W/m ² K, SHGC = 0.763; eight skylights in the core zone each measuring 1.5 m ² : U = 5.778 W/m ² K, SHGC = 0.819
Space conditioning	Fan-powered variable-air-volume system with an outside air economizer, hot water reheat coils, and central chilled water cooling coil
Infiltration	ACH ₅₀ = 0.1
Occupancy	One occupant for every 25.58 m ² of floor area
Internal Loads	Lighting = 10.8 W/m ² , Electrical Equipment = 16.7 W/m ²

Fig. 3 This model of an 800 m² office building has four exterior, one interior zone, and 40% window-to-wall ratio; for more details, the full model file is available (<http://bit.ly/24vxszC>).

5.2 Selection of Faults and Making a List of Tunable Parameters

Autotune requires a list of tunable parameters, which it alters to calibrate the model according to the measured data provided. This list is defined by the user based on how the model is to be used. For example, if only whole-building energy data are available, parameters involving the major contributors to whole-building energy use (e.g., HVAC equipment, water heating, and lighting) should be used. However, if all the measured data consists of only HVAC electricity consumption, then parameters that do not affect the HVAC system would not be required.

For this experiment, 24 types of faults, shown in Table 3, were injected into the input file by changing the original value of multiple parameters to another value. Since each of these 24 values may exist in multiple

objects in EnergyPlus, there were actually a total of 63 input file changes. The changes are too extensive to include here, but they are available electronically (bit.ly/1SQ12f2).

Table 3. The 24 types of faults used to inject 63 parameter changes into the original input file to provide a de-tuned file from which Autotune will attempt to recover the original.

Number	Class	Field
1	Material	Thickness
2	Material	Density
3	Material	Solar Absorptance
4	Material:NoMass	Thermal Resistance
5	Material:NoMass	Solar Absorptance
6	WindowMaterial:SimpleGlazingSystem	U-Factor
7	WindowMaterial:SimpleGlazingSystem	Solar Heat Gain Coefficient
8	People	Number of People
9	People	Fraction Radiant
10	Lights	Lighting Level
11	Lights	Return Air Fraction
12	Lights	Fraction Radiant
13	Lights	Fraction Visible
14	ElectricEquipment	Design Level
15	ElectricEquipment	Fraction Latent
16	ElectricEquipment	Fraction Radiant
17	ElectricEquipment	Fraction Lost
18	Exterior:Lights	Design Level
19	Fan:VariableVolume	Fan Efficiency
20	Fan:VariableVolume	Pressure Rise
21	Fan:VariableVolume	Motor Efficiency
22	Pump:VariableSpeed	Motor Efficiency
23	Boiler:HotWater	Nominal Thermal Efficiency
24	Chiller:Electric:EIR	Reference COP

The tunable parameters are defined in a *.csv file, which allows a user to define: a *class*, *object*, and *field* to uniquely identify any parameter in an EnergyPlus input file; *default*, *minimum*, and *maximum* for the starting and range of permitted variables; *distribution* of uncertainty (e.g., uniform or normal) for the parameter as supported by the SciPy library; *data type* (e.g., float or integer); *group* to define multiple parameters that function as one statistical degree of freedom (e.g., multiple zones may be treated as having the same, building-wide infiltration rate); and *constraint*, which allows mathematical combinations of groups to be performed that would otherwise result in a faulty simulation (e.g., the thermostat heating set point cannot be higher than the cooling set point).

To simplify settings for this experiment, all parameters were varied by 40% from the detuned value (e.g., a minimum of 60 and maximum of 140 for a parameter with a default value of 100) unless this pushed the variable beyond a physically realistic limit; in that case, the value was clamped. The final file of tunable input parameters is available electronically named (bit.ly/1SQ12f2).

5.3 Measured Data

Measured data can be provided in a *.csv file using with a column for each channel of data. The first column serves as the Date/Time stamp whereas subsequent columns contain a header row of a variable name that must correspond to an EnergyPlus output, and each row contains measured data from that timestep for all columns. In the common use case where individuals calibrate to electric utility bills, “WholeBuilding:Total Electric Demand” is used with a reporting frequency of monthly or hourly. In this experiment, we used simulation output from a simulation of the original, manually-calibrated input file as a surrogate for measured data.

5.4 Running the Calibration

Using the Autotune code (bit.ly/autotune_code) stand-alone demo folder, typing `autotune.py` at a command prompt reveals the command line interface options for running Autotune. The typical usage points Autotune to the files described in Section 3.2 as follows:

```
python autotune.py --schedule myschedule.csv --popsize 8 --maxevals 16 myidf.idf myparams.csv  
myuserdata.csv myweather.epw
```

Other options include

- `--numcpus`: number of cores permitted for Autotune calibration
- `--popsize`: number of individuals at each generation
- `--maxevals`: maximum number of fitness evaluations permitted
- `--maxtime`: maximum time allowed for a job (in seconds)

Once the files described in Section 4.1 – 4.3 are placed in the demo directory, `autotune.bat` can be used to start Autotune and provide performance statistics during the calibration process. When the Autotune calibration is complete, an output folder will store the statistics (`stat.csv`), the genome of the created buildings at each generation (`inds.csv`), and the final generation of calibrated building models (`final_[001-016].idf`).

5.5 Results

For this experiment, the calibrations were run with increasing values of `--maxevals` to 16, 32, 256, 512, and 1,024 to check the variation of this Autotune parameter on final accuracy of the calibrated models. These evaluations were organized into 16 individuals (building models) evolved over 64 generations, tournament selection with tournament size 4, generational replacement with weak elitism (one elite), heuristic crossover, and Gaussian mutation with a usage rate of 1.0 and a mutation rate of 10% of the allowable range of each variable, unless otherwise specified. Since this experiment uses simulation output as surrogate measured data, it allows reproducibility at the expense of realism. Specifically, sources of error such as model-form uncertainty (simplified algorithms not matching true physics), actual weather data for the building, and differences from as-modeled occupancy that would otherwise be issues are eliminated. The input file with the greatest fitness (i.e. `final_003_512`) is available electronically (bit.ly/1SQ12f2).

5.5.1 Fitness Evaluations

For the computing system used in this study, it took ~60 seconds to run an annual simulation of the test building. To determine how well a newly constructed building model's output matches measured data, the computationally expensive simulation was run for each building. Best results were obtained in Autotune run with `maxevals` values 512 and 1024 which took 112 and 220 minutes respectively.

5.5.2 Calibration Accuracy

The best ten files in terms of the lowest CV(RMSE) were selected from the final pool of 16 files generated by each of the Autotune calibrations with `maxevals` of 256, 512, and 1,024. MBE and CV(RMSE) values for these files are plotted in Figure 4; the best CV(RMSE) is 1.68% and the best MBE is 0.42%. In Figure 4, the filenames on the x-axis are named "final_xxx_yyy" where xxx is the file number of one of 16 candidate solution files generated by Autotune and yyy value of `maxevals` (e.g., 256, 512 or 1024).

Fig. 4. Comparison of MBE and CV(RMSE) for daily electric energy use for the ten best building models discovered from each of the 256, 512, and 1,024-evaluation Autotune calibration runs calibrating to hourly data. Best CV_RMSE was seen in *final_003_512*.

Although CV(RMSE) and MBE serve as summary statistics, it is often informative to plot individual data points before and after calibration to qualitatively assess calibration accuracy, as shown in Figure 5.

Fig. 5. Daily electric energy use is shown as a function of outdoor air temperature to demonstrate the accuracy of simulation before (left) and after (right) Autotune calibration. Autotune calibration reduced CV(RMSE)=31.04% and MBE=27.46% to CV(RMSE)=1.7% and MBE=0.4% (*final_003_512*).

In this paper, daily simulated and measured energy uses will be plotted as a function of daily outdoor air temperature as shown in Figure 5. As building's energy use has a strong correlation with weather related parameter such as outdoor air temperature or cooling/heating degree days [124-129], this type of plot can present normalized comparison between measured and corresponding simulated energy uses. Regression analysis is frequently used to define the relationship between the energy use and the outdoor temperature. Figure 6 shows an example of simple linear model that shows a daily heating energy use as a function of outdoor air temperature [130].

$$E = \beta_1 + \beta_2 T_o$$

where β_1 and β_2 are regression coefficients and T_o is the outside air dry-bulb temperature. Other works have reported relationship between energy use and outdoor temperature as non-linear models [131,132]. At an annual level, this is a common method by which experts iteratively calibrate buildings manually, by modifying loads so that building energy use for heating and cooling are reacting appropriately to environmental conditions [129, 133-136].

Fig. 6. A conceptual daily heating energy signature

5.5.3 Convergence

Each generation, by default, has 16 individuals/buildings used for the calibration. As the generations progress, we provide statistics on each set of 16 buildings regarding the error rate. This includes the worst (maximum), best (minimum), medium, average, and standard deviation of the error rate. As Autotune calibration progresses to subsequent generations, the error rates should decrease.

Figure 7 graphically displays the performance statistics saved during the calibration process. As can be seen, Autotune converged quickly to a set of good models but could make only small improvements over a long time to create the best-calibrated models.

Fig. 7. Autotune quickly finds a set of good models by reducing the error in terms of Watts (W), but takes longer to create the best models.

5.5.4 Recovery of Original Inputs

This experiment began with a manually-calibrated input file which was detuned by adding faults to 63 variables by randomly varying the value of each variable up to 40% from the original value. Autotune was never exposed to the original values for these parameters, and can thus be used to assess Autotune's ability to recover the original values for these input parameters in its best-calibrated model. In summary, 32 of the 63 (50%) were within 10% of their original value, 22 (35%) were within 30% of their original value, and the remaining 9 (14%) parameters had a difference above 30%. The variable with the lowest difference was one of the most important when it comes to energy use, the Reference COP of Chiller:Electric:EIR which was off by only 0.27%.

5.5.5 Comparison of HVAC Energy use

Energy to meet HVAC loads is almost always the largest source of energy consumption for a building. By considering such an important variable in more detail, we can determine how well the calibration process performed for data against which it could not directly compare. Figure 8 shows that by selecting the best-calibrated model (file_003_512) with CV(RMSE) = 5.65% and MBE = -4.57%, Autotune got relatively close to the daily HVAC energy use.

Fig. 8. Autotune closely matches energy consumption related to HVAC loads closely to the original model despite not using it as part of the calibration process.

5.5.6 Comparison of non-HVAC Energy use

The non-HVAC energy use portion of the whole building electric demand would include lights, electrical equipment, exterior lights, and equipment. Generally, these components use electrical power based solely upon the schedule of their use. This is unlike electrical loads from HVACs which could involve fans, electric coils, pumps, chillers, or cooling towers. Although Autotune was run against the Whole Building Electric demand, it is informative to assess how Autotune has performed in matching the non-HVAC energy use of the model to the original data.

The CV(RMSE) in non-HVAC energy use for this file was 6.93% and MBE was 6.52% as shown in Figure 9. This low error rate for both HVAC and non-HVAC energy use are understandably higher than whole building energy comparison since Autotune had no data to directly compare against for these individual types of equipment. The difference underscores the potential for sub-metering to enhance automatic calibration. It would be an informative study to determine which sets of virtually sub-metered data yield the most-improved calibration models.

Fig. 9. For non-HVAC energy use Autotune calibration was able to bring down the error to CV(RMSE) of 6.93% and MBE of 6.52%. Autotune is able to recover daily energy use within 100 kWh/day for data which was not used in the calibration process. The three data rows in the figure represent energy use on weekdays (top), Saturday (middle) and Sunday (bottom).

5.5.7 Comparison of zone temperatures

Beyond total building energy consumption, electrical consumption for HVAC loads, and energy use by non-HVAC loads - occupant comfort, as measured by indoor radiant air temperature, is another important variable to consider. Specifically, was the calibrated model predicting the same distribution of indoor air temperature as the original model before de-tuning? Here, we used the average radiant temperature of all zones in the building and compare this value in both models. For the annual simulation of hour-by-hour data, we found a difference of 0-0.1°C for 53.5% of the time, 0.1-0.5°C for 45.8% of the time, 0.5-1°C for 0.5% of the time (48 hours during the year), and 1-1.3°C for 0.2% of the time (only 17 hours during the year). The average indoor radiant temperature varied 0.116°C during the 8760 hours. The difference in Mean Radiant Temperature (MRT) with respect to percentage is shown in Figure 10. Figure 11 shows the Normal Distribution Curve of Difference in MRT with a mean of 0.116°C and approximate standard deviation of 0.109.

Fig. 10. Average indoor mean radiant temperature annual variation for all the zones in the original and calibrated model matched closely, showing the ability to recover hourly temperatures within an average of 0.116°C despite not being used in calibration and being more removed from the electrical data utilized.

Fig. 11. Normal distribution curve of error between the original and calibrated model for the largest thermal zone's mean radiant temperature has a standard deviation of 0.109 which can be used to place a fairly tight uncertainty bounds around Autotune calibration for this case study.

6. Case Study 2: Manually Tuned Model Experiment

The research for this case study is based on data collected in November 2012 from ORNL's two pairs of low-energy research houses with simulated occupancy in the Wolf Creek Subdivision in Oak Ridge, Tennessee (www.zebralliance.com). These houses were designed to maximize energy efficiency using new, ultra-high-efficiency ECMs. The first pair (designated WC1 and WC2) had 344 m² (3700 ft²) of conditioned area each with a basement; the second pair (WC3 and WC4) had 253 m² (2,720 ft²) of conditioned area each with crawlspace foundations.

An original model of the WC4 house was created and detailed manual calibration of the WC4 model to measured field data was performed by Im et al. [64]. In this case study, the original, uncalibrated model for WC4 was calibrated by Autotune. The results of the manual calibration were then compared with the calibration using Autotune.

6.1 About the Model

Dwelling WC4 is a residential building with a conventional ventilated attic and sealed and semi-conditioned crawlspace foundation. The air-conditioning system in this house is a 2-ton, multi-zone control, dual-capacity air source heat pump (ASHP) with a variable-speed blower, a Seasonal Energy Efficiency Ratio (SEER) rating of 18.4, and a Heating Seasonal Performance Factor (HSPF) of 9.1. A summary of the house characteristics is provided in Table 4. In the table various insulations are represented by R-value which is the unit of thermal resistance for a particular material or assembly of materials.

EnergyPlus v7.0 (2011) does not allow multi-zone control for a multi-speed heat pump system; therefore, a similar HVAC model and controls were utilized. Figure 12 shows the EnergyPlus model of WC4 [64].

Table 4. The residential building used in this study was an energy-efficient building with high insulation, advanced windows, low infiltration, LED lighting, and many advanced technologies being tested against similar technologies in a twin-home.

Characteristics	Description
Floor area	253 m ² (2720 ft ²)
Foundation	Sealed crawlspace, no floor joist insulation, R-10 insulation on interior side of the block wall, floor framing factor = 13%
Exterior walls	R-21 whole wall value, wall framing factor = 23%, solar absorptance = 0.76
Attic	Conventional unconditioned vented attic, R-50 ceiling insulation (U=0.019 W/m ² K), attic wall insulation R-15, fan-induced air changes per hour of up to 10
Roofing material	Composition shingles, solar reflectance = -0.26, roof deck R-4 EPS insulation
Windows	Triple pane. Windows facing southeast and southwest: U = 0.24 W/m ² K, SHGC = 0.50; windows facing northeast and northwest: U = 0.18 W/m ² K, SHGC = 0.22
Space conditioning	Single ASHP, cooling capacity = 25 kBTu/h(7.32 kW), SEER 18.4, heating capacity 22.6 kBTu/h(6.62 kW), HSPF 9.1
Infiltration	Air changes per hour at a pressure difference of 50 Pascals (ACH ₅₀) = 2.3, Specific Leakage Area(SLA) = 0.00017 in ² /in ²
Mechanical ventilation	30 CFM (Cubic Feet per Minute)
Duct location	Supply, interior; return, interior; R-5 insulation, supply area = 551 ft ² (51.19 m ²), return area = 102 ft ² (9.47 m ²), duct air leakage (to outside) = 8%.
Air handler location	Interior
Water heater	Electric 50 gal(189.2 l) capacity, usage = 60 gal/day (227 l/day), set temp = 120° F (49° C), add-on heat pump COP = 2.4
Lighting	100% LED

Fig. 12. This is an EnergyPlus model of a 253 m² (2,720 ft²) residential research home located in Oak Ridge, TN, USA. It is in the Wolf Creek subdivision and thus referred to as WC4.

6.2 Measured Utility Data and Weather Data

The 15-minute interval weather data used in this experiment were collected from a weather station located at the WC4 site. The daily, monthly, and annual end-use data—including fan, heating and cooling energy, and overall energy consumption—were obtained from the hourly sensor data.

6.3 Initial Utility Data Comparison

The simulation results for the base model showed that the annual whole-building energy consumption from the model was 10,968 kWh, 20% less than the measured data. The daily MBE and the CV(RMSE) were 20.5% and 25.93%, respectively whereas for monthly it was 20.5% and 22.37% respectively. This is relatively good agreement for an initial, uncalibrated model, but the MBE and CV(RMSE) indicate that the initial model needs to be calibrated [64]. The low initial error was possible since the model was developed with known details of building materials and construction, measured data, and emulated occupancy (i.e., no uncertainty in occupancy behavior). The comparison between the initial model and the measured data is shown in Figure 13 and Figure 14 for daily and monthly respectively.

Fig. 13. Daily energy use comparison (measured vs. simulated) for the WC4 emulated-occupancy research home shows a relatively poor agreement in daily energy use for the original, uncalibrated

model. The daily errors between the measured utility data and uncalibrated EnergyPlus were a CV(RMSE) of 25.93% and a MBE of 20.5%.

Fig. 14. Monthly energy use comparison (measured vs. simulated) for the WC4 emulated-occupancy research home shows the gap to be covered in the competition between manual versus automatic calibration for this case study. The daily errors between the measured utility data and uncalibrated EnergyPlus were a CV(RMSE) of 22.37% and a MBE of 20.5%.

The simulation output and measured data were compared on a daily basis to preserve the granularity and for better visual comparison, as the monthly data would not capture all the details and the hourly-level comparison would be too detailed and would present challenges for meaningful visual comparison.

6.4 Selection of Input Parameters

Autotune calibration was performed in a naïve fashion, with very limited information regarding the target building. Thirty-five types of generic parameters, which were believed to affect total electric demand, were selected for this calibration for a total of 166 tunable parameters. This is a very large number of calibration parameters compared to most calibration processes and is beyond what most humans and algorithms can handle. Lighting and plug power density variables were not included as they were set to known values in the original model. The emulated-occupancy equipment schedules allow these input parameters to be set to known/enforced values and reduce the computational complexity of additional parameters. Calibration parameters are typically defined in terms of which parameters are the most uncertain and which most affect energy use (or other performance variable such as zone temperature) for the use case relevant to the practitioner as the overall time of calibration increases with more parameters.

Material properties of the external walls, roof, and foundation were included since they can affect HVAC load. The types of tunable input parameters used in this Autotune calibration are listed in Table 5. Unlike the previous case study, maximum and minimum limits of input parameters were fixed based upon building scientists' assessment of physically realistic values for each parameter with the exception of variables of class Material and Ground Temperature which were varied by 30%.

Table 5. Thirty-five parameter types, involving 166 variable changes in the input file, were selected for Autotune calibration of the WC4 experimental house and constitute an extreme example of the number of input variables a manual calibration process would typically accommodate.

S.No.	Class	Field
1	Site:GroundTemperature:BuildingSurface	January Ground Temperature
2	Material	Thickness
3	Material	Density
4	Material	Thermal Absorptance
5	Material	Solar Absorptance
6	Material:NoMass	Thermal Resistance
7	Material:NoMass	Thermal Absorptance
8	Material:NoMass	Solar Absorptance
9	WindowMaterial:SimpleGlazingSystem	Solar Heat Gain Coefficient
10	WindowMaterial:SimpleGlazingSystem	Visible Transmittance
11	WindowMaterial:SimpleGlazingSystem	U-Factor
12	ZoneInfiltration:FlowCoefficient	Flow Coefficient
13	ZoneInfiltration:FlowCoefficient	Stack Coefficient
14	ZoneInfiltration:FlowCoefficient	Pressure Exponent
15	ZoneInfiltration:FlowCoefficient	Wind Coefficient
16	ZoneInfiltration:FlowCoefficient	Shelter Factor
17	Sizing:Zone	Heating Design Air Flow Rate
18	Fan:OnOff	Fan Efficiency
19	Fan:OnOff	Maximum Flow Rate
20	Fan:OnOff	Motor Efficiency
21	Coil:Cooling:DX:MultiSpeed	Crankcase Heater Capacity
22	Coil:Cooling:DX:MultiSpeed	Speed 1 Rated COP
23	Coil:Cooling:DX:MultiSpeed	Speed 2 Rated COP
24	Coil:WaterHeating:AirToWaterHeatPump	Rated Heating Capacity
25	Coil:WaterHeating:AirToWaterHeatPump	Rated COP
26	Coil:WaterHeating:AirToWaterHeatPump	Condenser Water Pump Power
27	Coil:WaterHeating:AirToWaterHeatPump	Fraction of Condenser Pump Heat to Water
28	Coil:WaterHeating:AirToWaterHeatPump	Crankcase Heater Capacity
29	WaterHeater:Mixed	Heater Thermal Efficiency
30	WaterHeater:Mixed	On Cycle Parasitic Fuel Consumption Rate
31	WaterHeater:Mixed	Off Cycle Parasitic Fuel Consumption Rate
32	WaterHeater:HeatPump	On Cycle Parasitic Electric Load
33	WaterHeater:HeatPump	Off Cycle Parasitic Electric Load
34	Schedule:Compact	Field 4
35	ElectricEquipment	Design Level

6.5 Autotune Calibration results: MBE and CV(RMSE) vs. Number of Fitness Evaluations

For the computing system used in this study, 114 seconds were required for an annual simulation of WC4. Similar to the previous experiment, Autotune was run with maxevals of 64, 128, 256, 512, 1,024 and 2,048. Autotune took 182, 400 and 720 minutes for maxevals value of 512, 1024 and 2048 respectively; the error rates for Autotune's five best-calibrated files in terms of lowest CV(RMSE) are shown in Figure 15. Autotune's best file (003_1024) had a CV(RMSE) = 11.82% and MBE = -1.27%. In Figure 15, the filenames on the x-axis are named "final_xxx_yyy" where xxx is the file number of one of 16 candidate solution files generated by Autotune and yyy value of maxevals (e.g., 256, 512 or 1024).

Fig. 15. Performance before and after Autotune calibration in terms of CV(RMSE) (left) and MBE (right) for the five best-calibrated models from 3 independent Autotune calibration trials. Original CV_RMSE and Original MBE are the error rates of the uncalibrated model. The gap between the red line and blue bars indicates the improvement achieved by Autotune calibration. Autotune shows the stability for these experiments by achieving similar error metrics in all independent trials.

6.6 Comparison with Manual Calibration

Since Autotune provides the final generation of calibrated building models (equal to the population size), we considered only the building with the lowest CV(RMSE) of 11.82% and MBE of -1.27% for comparison. The best CV(RMSE) and MBE achieved by Im et al. [64] after the ninth and final step of manual calibration were 10.4% and 0.18%, respectively. Figure 16 shows a comparison of the monthly energy consumption in the original EnergyPlus model, the manually calibrated model, the Autotune calibrated model, and the measured utility data. Figure 17 illustrates Autotune data vs. measured data, and Figure 18 shows manual calibration data vs. measured data.

Fig. 16. Final comparison of monthly energy consumption from the base model to the measured data by both the manual calibration and Autotune calibration efforts show a nearly identical ability in recovering the monthly data. Autotune calibration reduced the CV(RMSE) of 25.93% and the MBE of 20.5% to the CV(RMSE) of 11.82% and the MBE of -1.27%.

Fig. 17. Autotune calibration data vs. measured data for the WC4 emulated-occupancy research home. The daily errors between the measured utility data and Autotune calibrated EnergyPlus model were CV(RMSE) of 11.82% and MBE of -1.27%.

Fig. 18. Manual calibration data vs. measured data for the WC4 emulated-occupancy research home. The daily errors between the measured utility data and manually calibrated EnergyPlus model were CV(RMSE) of 10.4% and MBE of 0.18%.

6.7 Performance of Autotune

For Autotune run with *maxevals* value of 512, the best file of each 127 generations was extracted and compared with electrical data to show progress of Autotune calibration in each generation. Figure 19 shows the improvement of Autotune from the base case to the measured data at four points along its evolutionary calibration.

In Autotune, the building considered as the best calibrated among all the other individuals in a generation is the individual/building with a fitness of the best CV(RMSE) value (i.e. the minimum value). In each generation, the best individual of the previous generation becomes a parent which is compared with the fitness of all offspring from the current generation. If any offspring surpass the parent in fitness, it becomes the new best individual; otherwise the parent stays. For this reason, the same error rate occurs in a few consecutive generations in the later stages of calibration when no offspring surpassed the best parent. For

illustrative purposes in Figure 20, we instead plot the new/unique best individuals from each of the 127 generations in terms of CV(RMSE) and MBE.

Fig. 19. Monthly energy use calibration accuracy is shown as Autotune continuously improves its results at generations 1 (top-left), 10 (top-right), 50 (bottom-left), and 127 (bottom-right).

Fig. 20. CV(RMSE) and MBE error rate for monthly data of generation best invidual decreases over the time making Autotune calibration outputs more accurate with each generation cycle.

7. Conclusions and Discussions

In case study 2, Autotune and the manually calibrated model were comparable in terms of CV(RMSE) and MBE, with the manually calibrated model proving a slightly more accurate than the automated method. However, Autotune required much less calibration time. The manual calibration shown in Im et al. [64] was performed by two experienced researchers, and calibration took about 40–50 hours (not counting the time required to generate the base model). Autotune, using the same base case model, required 30–40 minutes to make the list of tunable input parameters and 400 minutes (6.7 hours) to calibrate via 1,024 simulation evaluations. The base model from which this experiment began was carefully developed with measured data from detailed information for building systems. The less accurate the base model, in general, the more time is required for the calibration. It is recommended that as much time as possible be spent on improving the base model before the calibration. This may be a necessity for the application of Autotune, since it can only modify what exists in the base model and cannot currently exercise the complete freedom of a manual calibrator to add or remove parts of the building models (e.g., objects in EnergyPlus).

This being a controlled experiment had several advantages as compared to a real building. Calibration of real building energy models will have several challenges with one being the inaccuracy of schedules which can alone cause whole-building electrical differences of up to 3 times annually and 5 times monthly between an energy saver and energy waster. It would be very difficult for a modeler to design the exact schedule for components in a building energy model without schedule information. Unfortunately, such levels of monitoring and data collection are not typically seen in buildings today. Tuning of schedule parameters, or of any time-series data, makes calibration computationally expensive and is an outstanding challenge in the field. Some automated calibrators mitigate this by treating schedules as a meta-parameter; this is done by defining schedule-related parameters according to an “average” schedule, an “energy saver” mode, and “energy waster” mode and then sampling/defining each schedule parameter for an individual based on a distribution around the average.

In case study 1, automated calibration is seen to match simulation results very well with the measured data, and Section 5.5.4 quantifies Autotune’s ability to recover 50% of detuned parameters to within a 10% of their original value from a 40% perturbed range for this test case. As shown in section 4.5.7, Autotune was successfully able to match mean radiant temperature profiles within the building without being exposed to this information during the calibration process. The MRT in the calibrated model was seen to vary by a maximum of only 0.5°C in 99.3% of hours in a year.

A major issue with BEPS models calibration is a lack of explicit standards/guidelines [49]. Current guidelines only specify broad ranges of allowable error in total energy consumption of a building, but none for uncertainty/inaccuracy in BEPS’ input parameters or zone-level environmental data inaccuracy such as uncertainty in zone mean radiant temperature profiles. A recent 20,000-building calibration study [124], identified that the industry-standard calibration to monthly energy use has less than a 40% correlation, in all cases, to successfully recovering an individual building’s input parameters. Particularly, the authors observe a need for enhanced sub metering and calibration requirements proceeding from the industry-standard “Calibration Level 1 (total electric demand)” to “Calibration Level 2 (HVAC, equipment, and lighting)”, and to “Calibration Level 3 (temperature profiles)”.

In summary, this paper explicitly demonstrated two case studies that could be used to assess figures of merit for calibration procedures: (1) recovery of a de-tuned model based on a comparison to data not used in the calibration, and (2) manual versus automatic calibration based on quantitative and qualitative assessments of the calibration accuracy. Where possible, the authors have endeavored to provide all data and details necessary to allow other calibration techniques to be compared using these two case studies.

8. Future Work

The core of the new standard involves reducing the input-side error and better predicting actual energy savings after a retrofit. The ability to predict energy savings is a practical use case of BEPS, and is generally much easier than successfully recovering all input values such that they match the actual building properties (regarded as a theoretical limit in calibration accuracy). Additional components such as the addition of noise to simulation outputs (i.e. emulate sensor error), or modification of inputs that are not calibrated (i.e. to capture real-world dynamics) may be added as the standard evolves and would serve to further extend the work presented in this study. All calibration methods to date were developed in an era when meeting ASHRAE Guideline 14 criteria was sufficient. New research is showing a correlation of less than 0.4 between output-side and input-side errors, as well as pointing to a need for more extensive sub-metering to use in

calibration. The next generation of calibration algorithms will require much more extensive experimentation and testing to create a calibrated building model that more closely matches real buildings. The new standard and two case studies presented (with all required supplementary data) should facilitate quantitative comparison of calibration technologies, maturation of current calibration tools, and enhance market viability.

9. Acknowledgment

Funding for this project was provided by field work proposal CEBT105 under Department of Energy Building Technology Activity Number BT0201000. This manuscript has been authored by UT-Battelle, LLC, under Contract Number DEAC05-00OR22725 with DOE. The United States Government retains and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this manuscript, or allow others to do so, for United States Government purposes.

References

- [1] Shnapp, S., & Lausten, J. (2013). Mitigation potential from India's buildings. Paris: Global buildings performance network, 10.
- [2] Energy Statistics 2013. National Statistical Organisation (NSO), Ministry of Statistics and Programme Implementation, Government of India, http://mospi.nic.in/mospi_new/upload/Energy_Statistics_2013.pdf, accessed 27 January 2016.
- [3] Energy Information Administration 2003. Commercial Buildings Energy Consumption Survey, <http://www.eia.gov/emeu/cbecs/contents.html>, accessed 27 January 2016.
- [4] Hong T, Lin H-W, Chang W-K, Yan D, Zhu D, Zhou X et al. Building performance simulation. Ernest Orlando Lawrence Berkeley National Laboratory Report. June 2013, LBNL-6418E.
- [5] Hong, T., Yang, L., Hill, D., & Feng, W. (2014). Data and analytics to inform energy retrofit of high performance buildings. *Applied Energy*, 126, 90-106.
- [6] Stuart, E., Goldman, C. A., Larsen, P., & Gilligan, D. (2014). The US ESCO industry: recent trends, current size and remaining market potential. Lawrence Berkeley National Laboratory, American Council for an Energy-Efficient Economy Proceedings, Chicago, Illinois, < <http://www.aceee.org/files/proceedings/2014/data/papers/3-319.pdf>> (November 30, 2014).
- [7] Stuart, E., Larsen, P. H., Goldman, C. A., & Gilligan, D. (2013). Current Size and Remaining Market Potential of US ESCO Industry. LBNL Report.
- [8] O'Neill, Z., Pang, X., Shashanka, M., Haves, P., & Bailey, T. (2014). Model-based real-time whole building energy performance monitoring and diagnostics. *Journal of Building Performance Simulation*, 7(2), 83-99.
- [9] Fulton, M., Baker, J., Brandenburg, M., Herbst, R., Cleveland, J., Rogers, J., & Onyeagoro, C. (2012). United States Building Energy Efficiency Retrofits: Market Sizing and Financing Models. New York: DB Climate Change Advisors and the Rockefeller Foundation.
- [10] Davito, B., Tai, H., & Uhlaner, R. (2010). The smart grid and the promise of demand-side management. *McKinsey on Smart Grid*, 3, 8-44.
- [11] Široký, J., Oldewurtel, F., Cigler, J., & Privara, S. (2011). Experimental analysis of model predictive control for an energy efficient building heating system. *Applied Energy*, 88(9), 3079-3087.
- [12] Pisello, A. L., Goretti, M., & Cotana, F. (2012). A method for assessing buildings' energy efficiency by dynamic simulation and experimental activity. *Applied Energy*, 97, 419-429.
- [13] Huang, Y., Niu, J. L., & Chung, T. M. (2013). Study on performance of energy-efficient retrofitting measures on commercial building external walls in cooling-dominant cities. *Applied Energy*, 103, 97-108.
- [14] Yang, Z., & Becerik-Gerber, B. (2015). A model calibration framework for simultaneous multi-level building energy simulation. *Applied Energy*, 149, 415-431.
- [15] Langner, M. R., Henze, G. P., Corbin, C. D., & Brandemuehl, M. J. (2012). An investigation of design parameters that affect commercial high-rise office building energy consumption and demand. *Journal of Building Performance Simulation*, 5(5), 313-328.
- [16] Coakley, D., Raftery, P., & Molloy, P. (2012, September). Calibration of whole building energy simulation models: Detailed case study of a naturally ventilated building using hourly measured data. In *First Building Simulation and Optimization Conference*, Loughborough, UK (pp. 10-11).
- [17] Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., ... & Tarantola, S. (2008). Global sensitivity analysis: the primer. John Wiley & Sons.
- [18] Akbari, H. (1995). Validation of an algorithm to disaggregate whole-building hourly electrical load into end uses. *Energy*, 20(12), 1291-1301.
- [19] Ghiass, C. (2006). Experimental estimation of building energy performance by robust regression. *Energy and buildings*, 38(6), 582-587.

- [20] Manfren, M., Aste, N., & Moshksar, R. (2013). Calibration and uncertainty analysis for computer models—a meta-model based approach for integrated building energy simulation. *Applied energy*, 103, 627-641.
- [21] Azadeh, A. G. H. A. D. E. R. I., Ghaderi, S. F., & Sohrabkhani, S. (2008). Annual electricity consumption forecasting by neural network in high energy consuming industrial sectors. *Energy Conversion and Management*, 49(8), 2272-2278.
- [22] Ekici, B. B., & Aksoy, U. T. (2009). Prediction of building energy consumption by using artificial neural networks. *Advances in Engineering Software*, 40(5), 356-362.
- [23] Magnier, L., & Haghighat, F. (2010). Multiobjective optimization of building design using TRNSYS simulations, genetic algorithm, and Artificial Neural Network. *Building and Environment*, 45(3), 739-746.
- [24] Chen, B. J., Chang, M. W., & Lin, C. J. (2004). Load forecasting using support vector machines: A study on EUNITE competition 2001. *Power Systems, IEEE Transactions on*, 19(4), 1821-1830.
- [25] Dong, B., Cao, C., & Lee, S. E. (2005). Applying support vector machines to predict building energy consumption in tropical region. *Energy and Buildings*, 37(5), 545-553.
- [26] Li, Q., Meng, Q., Cai, J., Yoshino, H., & Mochida, A. (2009). Applying support vector machine to predict hourly cooling load in the building. *Applied Energy*, 86(10), 2249-2256.
- [27] Wehenkel, L., & Pavella, M. (1993). Decision tree approach to power systems security assessment. *International Journal of Electrical Power & Energy Systems*, 15(1), 13-36.
- [28] Yu, Z., Haghighat, F., Fung, B. C., & Yoshino, H. (2010). A decision tree method for building energy demand modeling. *Energy and Buildings*, 42(10), 1637-1646.
- [29] Dhar, A., Reddy, T. A., & Claridge, D. E. (1999). A Fourier series model to predict hourly heating and cooling energy use in commercial buildings with outdoor temperature as the only weather variable. *Journal of solar energy engineering*, 121(1), 47-53.
- [30] Dhar, A., Reddy, T. A., & Claridge, D. E. (1999). Generalization of the Fourier series approach to model hourly energy use in commercial buildings. *Journal of solar energy engineering*, 121(1), 54-62.
- [31] Ji, Y., Xu, P., & Ye, Y. (2015). HVAC terminal hourly end-use disaggregation in commercial buildings with Fourier series model. *Energy and Buildings*, 97, 33-46.
- [32] Zhou, Y. P., Wu, J. Y., Wang, R. Z., Shiochi, S., & Li, Y. M. (2008). Simulation and experimental validation of the variable-refrigerant-volume (VRV) air-conditioning system in EnergyPlus. *Energy and buildings*, 40(6), 1041-1047.
- [33] Al-ajmi, F. F., & Hanby, V. I. (2008). Simulation of energy consumption for Kuwaiti domestic buildings. *Energy and buildings*, 40(6), 1101-1109.
- [34] Winkelmann, F. C., Birdsall, B. E., Buhl, W. F., Ellington, K. L., Erdem, A. E., Hirsch, J. J., & Gates, S. (1993). DOE-2 supplement: version 2.1 E (No. LBL--34947). Lawrence Berkeley Lab., CA (United States); Hirsch (James J.) and Associates, Camarillo, CA (United States).
- [35] ESRU. ESP-r. [accessed 26.03.16.]. <http://www.esru.strath.ac.uk/>.
- [36] USDE. 2009a. "EnergyPlus testing and validation, Department Energy, USA." [accessed 26.03.16]. <http://apps1.eere.energy.gov/buildings/energyplus>.
- [37] Solar Energy Laboratory. TRNSYS version 17.02 user manual and documentation, solar energy laboratory, mechanical engineering department. Madison, WI: University of Wisconsin; May 2014.
- [38] IEA ECBCS Annex 53, Annex 53 total energy use in buildings: analysis & evaluation methods, www.ecbcs53.org.
- [39] Energy Efficiency Verification Specialists, International Performance Measurement and Verification Protocol (IPMVP), 2002.
- [40] ASHRAE. Guideline 14-2002: measurement of energy and demand savings. Atlanta, GA 30329: American Society of Heating, Refrigerating and Air-Conditioning Engineers; 2002.
- [41] Hensen, J. L., & Lamberts, R. (Eds.). (2012). *Building performance simulation for design and operation*. Routledge.
- [42] Ji, Y., & Xu, P. (2015). A bottom-up and procedural calibration method for building energy simulation models based on hourly electricity submetering data. *Energy*, 93, 2337-2350.
- [43] Crawley, D. B., Hand, J. W., Kummert, M., & Griffith, B. T. (2008). Contrasting the capabilities of building energy performance simulation programs. *Building and environment*, 43(4), 661-673.
- [44] Crawley, D. B., Lawrie, L. K., Winkelmann, F. C., Buhl, W. F., Huang, Y. J., Pedersen, C. O., ... & Glazer, J. (2001). EnergyPlus: creating a new-generation building energy simulation program. *Energy and buildings*, 33(4), 319-331.
- [45] Fumo, N., Mago, P., & Luck, R. (2010). Methodology to estimate building energy consumption using EnergyPlus Benchmark Models. *Energy and Buildings*, 42(12), 2331-2337.
- [46] Pang, X., Bhattacharya, P., O'Neill, Z., Haves, P., Wetter, M., & Bailey, T. (2011, November). Real-time building energy simulation using EnergyPlus and the Building Controls Virtual Test Bed. In *Proceeding of the 12th IBPSA Conference* (pp. 2890-2896).
- [47] ASHRAE, A. G. (2002). *Guideline 14-2002: Measurement of Energy and Demand Savings*. American Society of Heating, Refrigerating and Air-Conditioning Engineers, Atlanta.
- [48] New, J., & Chandler, T. (2013). Evolutionary tuning of building models to monthly electrical consumption. *ASHRAE Transactions*, 119, 89.

- [49] Tupper, K., Franconi, E., Chan, C., Fluhrer, C., Jenkins, M., & Hodgin, S. (2011). Pre-Read for BEM innovation summit. Retrieved from Rocky Mountain Institute website: [http://www.rmi.org/Content/Files/Summit_PreRead_Apr-19-2011\(2\).pdf](http://www.rmi.org/Content/Files/Summit_PreRead_Apr-19-2011(2).pdf).
- [50] Coakley, D., Raftery, P., & Keane, M. (2014). A review of methods to match building energy simulation models to measured data. *Renewable and Sustainable Energy Reviews*, 37, 123-141.
- [51] Raftery, P., Keane, M., & Costa, A. (2011). Calibrating whole building energy models: Detailed case study using hourly measured data. *Energy and Buildings*, 43(12), 3666-3679.
- [52] Mustafaraj, G., Marini, D., Costa, A., & Keane, M. (2014). Model calibration for building energy efficiency simulation. *Applied Energy*, 130, 72-85.
- [53] Karlsson, F., Rohdin, P., & Persson, M. L. (2007). Measured and predicted energy demand of a low energy building: important aspects when using building energy simulation. *Building services engineering research and technology*, 28(3), 223-235.
- [54] Turner C, Frankel M. *Energy performance of LEED for new construction buildings*. Washington, DC: New Buildings Institute; 2008.
- [55] Scofield, J. H. (2009). Do LEED-certified buildings save energy? Not really.... *Energy and Buildings*, 41(12), 1386-1390.
- [56] Ahmad, M., & Culp, C. H. (2006). Uncalibrated building energy simulation modeling results. *HVAC&R Research*, 12(4), 1141-1155.
- [57] Feng, X., Yan, D., & Hong, T. (2015). Simulation of occupancy in buildings. *Energy and Buildings*, 87, 348-359.
- [58] Pan, Y., Huang, Z., Wu, G., & Chen, C. (2006). The application of building energy simulation and calibration in two high-rise commercial buildings in Shanghai. *Proceedings of SimBuild*, 2-4.
- [59] Oldewurtel, F., Sturzenegger, D., & Morari, M. (2013). Importance of occupancy information for building climate control. *Applied Energy*, 101, 521-532.
- [60] Gugliemetti, F., Passerini, G., & Bisegna, F. (2004). Climate models for the assessment of office buildings energy performance. *Building and Environment*, 39(1), 39-50.
- [61] Hong, T., Chang, W. K., & Lin, H. W. (2013). A fresh look at weather impact on peak electricity demand and energy use of buildings using 30-year actual weather data. *Applied Energy*, 111, 333-350.
- [62] Menezes, A. C., Cripps, A., Bouchlaghem, D., & Buswell, R. (2012). Predicted vs. actual energy performance of non-domestic buildings: Using post-occupancy evaluation data to reduce the performance gap. *Applied Energy*, 97, 355-364.
- [63] Li, C., Hong, T., & Yan, D. (2014). An insight into actual energy use and its drivers in high-performance buildings. *Applied Energy*, 131, 394-410.
- [64] Im P, Bhandari M. Is monthly and whole building level calibration enough?: A detailed modeling and calibration study of an ultra-efficient occupancy simulated house. In: *ASHRAE/IPBSA-USA building simulation conference*; 2014. p. 177-86.
- [65] Hong, T., Yang, L., Hill, D., & Feng, W. (2014). Data and analytics to inform energy retrofit of high performance buildings. *Applied Energy*, 126, 90-106.
- [66] Clarke, J. A. (1993). Assessing building performance by simulation. *Building and Environment*, 28(4), 419-427.
- [67] Sonderegger R, Avina J, Kennedy J, Bailey P. Deriving loadshapes from utility bills through scaled simulation. ASHRAE seminar presentation, Kansas City, MI; 2001.
- [68] Yin, R., Kiliccote, S., & Piette, M. A. (2015). Linking measurements and models in commercial buildings: A case study for model calibration and demand response strategy evaluation. *Energy and Buildings*.
- [69] Pedrini, A., Westphal, F. S., & Lamberts, R. (2002). A methodology for building energy modelling and calibration in warm climates. *Building and Environment*, 37(8), 903-912.
- [70] Clarke, J. A., Strachan, P. A., & Pernot, C. (1993). An approach to the calibration of building energy simulation models. *TRANSACTIONS-AMERICAN SOCIETY OF HEATING REFRIGERATING AND AIR CONDITIONING ENGINEERS*, 99, 917-917.
- [71] Tahmasebi, F., & Mahdavi, A. (2013, August). A two-staged simulation model calibration approach to virtual sensors for building performance data. In *Proceedings of the 13th Conference of International Building Performance Simulation Association*, Chambéry, France (pp. 25-28).
- [72] Diamond, S. C., & Hunn, B. D. (1981). Comparison of DOE-2 computer program simulations to metered data for seven commercial buildings (No. LA-UR-80-2791; CONF-810101-2). Los Alamos Scientific Lab., NM (USA).
- [73] Kaplan, M. B., McFerran, J., Jansen, J., & Pratt, R. (1990). Reconciliation of a DOE2. 1C model with monitored end-use data for a small office building. *ASHRAE transactions*, 96(1), 981-993.
- [74] Bronson, D. J., Hinchey, S. B., Haberl, J. S., & O'Neal, D. L. (1992). Procedure for calibrating the DOE-2 simulation program to non-weather-dependent measured loads. In *ASHRAE Winter Meeting*, Anaheim, CA, USA, 01/25-29/92 (pp. 636-652).
- [75] Bou-Saada TE, Haberl JS. An improved procedure for developing calibrated hourly simulation models. In: *Proceedings of the 5th international IBPSA conference*. Montréal, Canada; 1995.
- [76] Kreider, J. F., & Haberl, J. S. (1994). Predicting hourly building energy usage. *ASHRAE Journal* (American Society of Heating, Refrigerating and Air-Conditioning Engineers);(United States), 36(6).

- [77] Kreider, J. F., & Haberl, J. S. (1994). Predicting hourly building energy use: The great energy predictor shootout-- Overview and discussion of results (No. CONF-9406105--). American Society of Heating, Refrigerating and Air-Conditioning Engineers, Inc., Atlanta, GA (United States).
- [78] US DOE. M&V guidelines: measurement and verification for federal energy projects version 3.0. US Department Of Energy; 2008.
- [79] Reddy, T. A. (2006). Literature Review on Calibration of Building Energy Simulation Programs: Uses, Problems, Procedures, Uncertainty, and Tools. ASHRAE transactions, 112(1).
- [80] Fumo, N. (2014). A review on the basics of building energy estimation. Renewable and Sustainable Energy Reviews, 31, 53-60.
- [81] Sanyal, J., & New, J. (2013, May). Simulation and big data challenges in tuning building energy models. In Modeling and Simulation of Cyber-Physical Energy Systems (MSCPES), 2013 Workshop on (pp. 1-6). IEEE.
- [82] Waltz J. Practical experience in achieving high levels of accuracy in energy simulations of existing buildings. ASHRAE Trans 1992; 98(Part 1): 606–17 [paper number AN-92-1-2].
- [83] Haberl, J. S., O'Neal, D. L., Bronson, J. D., & Hinchey, S. B. (1993). Graphical tools to help calibrate the DOE-2 simulation program. ASHRAE Journal (American Society of Heating, Refrigerating and Air-Conditioning Engineers);(United States), 35(1).
- [84] Haberl, J. S., & Abbas, M. (1998). Development of graphical indices for viewing building energy data: Part I. Journal of solar energy engineering, 120(3), 156-161.
- [85] Haberl, J. S., & Abbas, M. (1998). Development of graphical indices for viewing building energy data: Part II. Journal of solar energy engineering, 120(3), 162-167.
- [86] Sun, K., Yan, D., Hong, T., & Guo, S. (2014). Stochastic modeling of overtime occupancy and its application in building energy simulation and calibration. Building and Environment, 79, 1-12.
- [87] Liu, M., & Claridge, D. E. (1998). Use of calibrated HVAC system models to optimize system operation. Journal of solar energy engineering, 120(2).
- [88] Wei G,Liu M,Claridge D.Signatures of heating and cooling energy consumption for typical AHUs. In: Proceedings of the eleventh symposium on improving building systems in hot and humid climates. Forth Worth, TX; 1998. p. 387–402.
- [89] Yu, F. W., & Chan, K. T. (2005). Energy signatures for assessing the energy performance of chillers. Energy and buildings, 37(7), 739-746.
- [90] Liu, M., Wei, G., & Claridge, D. E. (1998). Calibrating AHU models using whole building cooling and heating energy consumption data. Proc. of ACEEE 1998 Summer Study on Energy Efficiency in Buildings, 3, 229-241.
- [91] Liu, G., & Liu, M. (2011). A rapid calibration procedure and case study for simplified simulation models of commonly used HVAC systems. Building and Environment, 46(2), 409-420.
- [92] Monfet, D., Charneux, R., Zmeureanu, R., & Lemire, N. (2009). Calibration of a Building Energy Model Using Measured Data. ASHRAE Transactions, 115(1).
- [93] Parker, J., Cropper, P., & Shao, L. (2012). A calibrated whole building simulation approach to assessing retrofit options for Birmingham airport. Proceedings of the IBPSA—England.
- [94] Yoon, J., Lee, E. J., & Claridge, D. E. (2003). Calibration procedure for energy performance simulation of a commercial building. Journal of solar energy engineering, 125(3), 251-257.
- [95] Yoon J, Lee E. Calibration procedure of energy performance simulation model for a commercial building. In: Proceedings of building simulations 1999. Kyoto, Japan: International Building Performance Simulation Association; 1999.
- [96] Coakley, D., Raftery, P., Molloy, P., & White, G. (2011, November). Calibration of a detailed BES model to measured data using an evidence-based analytical optimisation approach. In Proceedings of the 12th International IBPSA Conference. Sydney, Australia.
- [97] Westphal, F. S., & Lamberts, R. (2005, August). Building simulation calibration using sensitivity analysis. In Ninth International IBPSA Conference (pp. 1331-1338).
- [98] Pan, Y., Huang, Z., & Wu, G. (2007). Calibrated building energy simulation and its application in a high-rise commercial building in Shanghai. Energy and Buildings, 39(6), 651-657.
- [99] Raftery, P., Keane, M., & O'Hagan, A. (2001). Calibrating whole building energy models: An evidence-based methodology. Energy and Buildings, 43(9), 2356-2364.
- [100] Carroll, W. L., & Hitchcock, R. J. (1993). Tuning simulated building descriptions to match actual utility data: methods and implementation. ASHRAE Transactions-American Society of Heating Refrigerating Airconditioning Engin, 99(2), 928-934.
- [101] Reddy, T. A., Maor, I., & Panjapornpon, C. (2007). Calibrating detailed building energy simulation programs with measured data—Part I: General methodology (RP-1051). Hvac&R Research, 13(2), 221-241.
- [102] Sun, J., & Reddy, T. A. (2006). Calibration of building energy simulation programs using the analytic optimization approach (RP-1051). HVAC&R Research, 12(1), 177-196.
- [103] Kennedy, M. C., & O'Hagan, A. (2001). Bayesian calibration of computer models. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 63(3), 425-464.
- [104] MacKay, D. J. (1994). Bayesian nonlinear modeling for the prediction competition. ASHRAE transactions, 100(2), 1053-1062.

- [105] Lomas, K. J., & Eppel, H. (1992). Sensitivity analysis techniques for building thermal simulation programs. *Energy and buildings*, 19(1), 21-44.
- [106] Heo, Y., Choudhary, R., & Augenbroe, G. A. (2012). Calibration of building energy models for retrofit analysis under uncertainty. *Energy and Buildings*, 47, 550-560.
- [107] O'Neill, Z., & Eisenhower, B. (2013, December). Leveraging the analysis of parametric uncertainty for building energy model calibration. In *Building simulation* (Vol. 6, No. 4, pp. 365-377). Springer Berlin Heidelberg.
- [108] Robertson, J. J., Polly, B. J., & Collis, J. M. (2015). Reduced-order modeling and simulated annealing optimization for efficient residential building utility bill calibration. *Applied Energy*, 148, 169-177.
- [109] O'Neill, Z., & O'Neill, C. (2016). Development of a probabilistic graphical model for predicting building energy performance. *Applied Energy*, 164, 650-658.
- [110] Li, Q., Gu, L., Augenbroe, G., Wu, C. J., & Brown, J. (2015). Calibration of Dynamic Building Energy Models with Multiple Responses Using Bayesian Inference and Linear Regression Models. *Energy Procedia*, 78, 979-984.
- [111] Tian, W., Yang, S., Li, Z., Wei, S., Pan, W., & Liu, Y. (2016). Identifying informative energy data in Bayesian calibration of building energy models. *Energy and Buildings*, 119, 363-376.
- [112] Ruiz, G. R., Bandera, C. F., Temes, T. G. A., & Gutierrez, A. S. O. (2016). Genetic algorithm for building envelope calibration. *Applied Energy*, 168, 691-705.
- [113] New, J. R., Sanyal, J., Bhandari, M., & Shrestha, S. (2012). Autotune E+ building energy models. *Proceedings of the 5th National SimBuild of IBPSA-USA*, 1-3.
- [114] Sanyal, J., New, J., & Edwards, R. (2013, July). Supercomputer assisted generation of machine learning agents for the calibration of building energy models. In *Proceedings of the Conference on Extreme Science and Engineering Discovery Environment: Gateway to Discovery* (p. 22). ACM.
- [115] Garrett, A., & New, J. (2015). Scalable tuning of building models to hourly data. *Energy*, 84, 493-502.
- [116] Dong, B., O'Neill, Z., Luo, D., & Trevor, B. (2013). Development and calibration of a reduced-order energy Performance model for a mixed-use building. In *Proceedings of Building Simulation*.
- [117] Kim, D., & Braun, J. E. (2015). A general approach for generating reduced-order models for large multi-zone buildings. *Journal of Building Performance Simulation*, 8(6), 435-448.
- [118] Michalewicz, Z., & Fogel, D. B. (2013). *How to solve it: modern heuristics*. Springer Science & Business Media.
- [119] De Jong, K. A., & Spears, W. M. (1993, June). On the State of Evolutionary Computation. In *ICGA* (pp. 618-625).
- [120] Spears, W. M., De Jong, K. A., Bäck, T., Fogel, D. B., & De Garis, H. (1993, April). An overview of evolutionary computation. In *Machine Learning: ECML-93* (pp. 442-459). Springer Berlin Heidelberg.
- [121] Fogel, D. B. (1994). An introduction to simulated evolutionary optimization. *Neural Networks, IEEE Transactions on*, 5(1), 3-14.
- [122] Fogel, D. B. (2000). What is evolutionary computation?. *Spectrum, IEEE*, 37(2), 26-28.
- [123] De Jong, K. A. (2006). *Evolutionary computation: a unified approach*. MIT press.
- [124] M.D. Lyberg. Source book for energy auditors. International Energy Agency, Energy Conservation in Buildings and Community Systems Programme, Annex XI, Energy Auditing, Vol. 1, pp. 38-42; Swedish Council for Building Research, Stockholm (1987).
- [125] Loftness, V., & Crenshaw, R. (1985). The refinement of descriptive indices critical to improving the thermal performance of whole buildings. In *Proceedings of the ASHRAE/DOE/BTECC Conference Thermal Performance of the Exterior Envelopes of Buildings III*, American Society of Heating, Refrigerating and Air-Conditioning Engineers, Atlanta, GA (pp. 289-295).
- [126] Burch, D. M., Cavanaugh, K., & Licitra, B. A. (1985). Effect of interior mass surfaces on the space heating and cooling load of a single-family residence. In *Proceedings of the ASHRAE/DOE/BTECC Conference Thermal Performance of the Exterior Envelopes of Buildings III*, American Society of Heating, Refrigerating and Air-Conditioning Engineers, Atlanta, GA (pp. 239-254).
- [127] Jacobsen, F. R. (1985). Energy signature and energy monitoring in building energy management systems. In *Proceeding of CLIMA 2000 World Congress* (Vol. 3, pp. 25-31).
- [128] Deeble, V. C., & Probert, S. D. (1986). Straight-line correlations for annual energy-consumption predictions?. *Applied energy*, 25(1), 23-39.
- [129] Kissock, J. K., Reddy, T. A., & Claridge, D. E. (1998). Ambient-temperature regression analysis for estimating retrofit savings in commercial buildings. *Journal of Solar Energy Engineering*, 120(3), 168-176.
- [130] Zmeureanu, R. (1990). Assessment of the energy savings due to the building retrofit. *Building and Environment*, 25(2), 95-103.
- [131] Fels, M. F. (1986). PRISM: an introduction. *Energy and Buildings*, 9(1-2), 5-18.
- [132] Palmiter, L. S., & Hanford, J. W. (1986). Relationship between electrical loads and ambient temperature in two monitored commercial buildings. *ASHRAE Trans.:(United States)*, 92(CONF-8606125-).
- [133] Chong, H. (2012). Building vintage and electricity use: Old homes use less electricity in hot weather. *European Economic Review*, 56(5), 906-930.
- [134] Yoon, J., Lee, E. J., & Claridge, D. E. (2003). Calibration procedure for energy performance simulation of a commercial building. *Journal of solar energy engineering*, 125(3), 251-257.

- [135] Kim, K. H., & Haberl, J. S. (2015). Development of methodology for calibrated simulation in single-family residential buildings using three-parameter change-point regression model. *Energy and Buildings*, 99, 140-152.
- [136] Liu, M., Song, L., Wei, G., & Claridge, D. E. (2003, January). Simplified building and air handling unit model calibration and applications. In *ASME 2003 International Solar Energy Conference* (pp. 15-25). American Society of Mechanical Engineers.
- [137] Garrett, A., & New, J. R. (2016, January). Suitability of ASHRAE Guideline 14 Metrics for Calibration. Technical paper. In *Proceedings of the ASHRAE Winter Conference and ASHRAE Transactions*, 122(1).