Applied Energy 124 (2014) 325–334

Contents lists available at ScienceDirect

Applied Energy

journal homepage: www.elsevier.com/locate/apenergy

Methods for benchmarking building energy consumption against its past or intended performance: An overview

Zhengwei Li *, Yanmin Han, Peng Xu

College of Mechanical Engineering, Tongji University, Shanghai, China
Institute of Advanced Study, Tongji University, Shanghai, China

HIGHLIGHTS

- This paper reviews up to date methods for building energy benchmarking.
- This paper summarizes the major characteristics of these methods.
- This paper recommends a flow chart for reader to choose a proper method.

ARTICLE INFO

Article history:
Received 10 December 2013
Received in revised form 6 March 2014
Accepted 14 March 2014
Available online 4 April 2014

Keywords:
Building
Energy benchmarking
Overview

ABSTRACT

Building sector consumes a significant portion of energy worldwide. One of the reasons is that the performance of building and its components degrades over the years. It is found that by improving the performance of existing systems through continuous commissioning, significant energy saving can be achieved. In a continuous commissioning process, energy benchmarking is extremely important for tracking, monitoring and detecting abnormal energy consumption behavior of a building. In this paper, up to date methods and tools available for energy benchmarking purpose are reviewed. It is hoped that with this paper, researchers and building operators are more confident in choosing a proper method (or tool) during the commissioning process.

© 2014 Elsevier Ltd. All rights reserved.

Contents

* Corresponding author at: College of Mechanical Engineering, Tongji University, Shanghai, China. Tel.: +86 021 65989750.
E-mail address: zhengwei_li@tongji.edu.cn (Z. Li).

http://dx.doi.org/10.1016/j.apenergy.2014.03.020
0306-2619/© 2014 Elsevier Ltd. All rights reserved.

1. Introduction

Building sector consumes more than 30% of the total energy worldwide [1]. An efficient way to alleviate global warming and improve environmental sustainability is to enhance building energy efficiency. However, many causes lead to a decrease of building energy efficiency for most buildings over the years, such as degradation of HVAC system components [2].

Continuous commissioning (CC) is an emerging technology to improve energy efficiency. Essentially, a CC is to conduct commissioning continuously throughout the life cycle of a building. It aims at assessing, improving and optimizing the performance of building systems [3]. A CC project launched in Texas A&M University between 1995 and 2000 is estimated to bring a total cost saving of as much as $10 million [4]. According to the Federal Energy Management Program (FEMP) in United States, CC has produced typical savings of 20% with payback under three years (often one to two years) in more than 130 large buildings [5]. To assess the energy consumption performance of a building, energy benchmarking is a necessary step. Energy benchmarking is defined as 'a macroscopic level of performance assessment, using metrics to measure its performance relative to other building or its previous performance' [6].

In the past, various methods for energy benchmarking have been developed. These methods can be categorized into white box method, gray box method and black box method. A white box method is also termed as first principle based method, which embeds physical constraints into the modelling of building components, and thus requires large amount of design documentations. Examples of this type of method include modified bin method and detailed energy simulation method [7,8]. On the contrast, a black box method uses data fitting techniques rather than physical knowledge, therefore requires a pre-selected statistical model and training data. Examples of black box method include artificial neural network method (ANN) and support vector machine method (SVM) [9]. The principle of gray box method lies in the middle between white box method and black box method, it combines both physical knowledge of the system and data fitting techniques to derive a useful energy model. Degree day method and its variants are examples of gray box method [10].

Benchmarking methods can also be categorized based on their corresponding types of baselines. Four types of baselines can be calculated by existing benchmarking methods: previous performance of comparable buildings, current performance of comparable buildings, previous performance of the same building, and intended performance of the same building [11]. While the first two types of baselines are often used by regulators and released to public, to encourage owners to improve energy efficiencies of their buildings [12], the rest are often used internally for energy tracking and monitoring purpose. In the context of this paper, benchmarking methods to calculate the latter two types of baselines are focused.

Overviews of building energy benchmarking methods have been given by several researchers [9,10,13]. Al-Homoud [10] introduced characteristics of mainly three methods: degree day based method, modified bin method, and detailed energy simulation method, which lie in the white box and gray box category. Holcomb [9] compared performance of three black box methods: multiple linear regression (MLR), artificial neural network (ANN) and support vector machine (SVM). It is found that ANN has the worst prediction accuracy compared to the other two methods. Zhao [13] investigated major characteristics of engineering methods (namely white box methods in this paper), statistical methods (namely regression based black box methods in this paper), neural networks, support vector machines, and grey models, with a conclusion that the neural network method and simplified engineering method have the highest accuracy.

To have a systematic view of up to date energy benchmarking methods and their performance levels, a literature review is conducted in this paper. It is hoped that with this paper, researchers can choose an appropriate benchmarking method based on the detail level of available information and required prediction accuracy. The content is organized as the following: first, the principles and characteristics of various benchmarking methods are introduced; second, the application cases of these methods and their performances are presented; finally, discussion section and conclusion remarks are given.

2. Energy benchmarking methods

As mentioned above, current energy benchmarking methods can be categorized into black box method, gray box method, and white box method. In this section, main methods in each category are briefly reviewed.

2.1. Black box method

In the category of black box method, multiple linear regression (MLR), bin method (BM), support vector regression (SVR), artificial neural network (ANN), and Gaussian process regression (GPR) are the most popular methods for energy benchmarking purposes.

2.1.1. Bin method (BM)

In this method, historical loads are grouped together into a bin if their associated variables (such as hour of week, temperature, and humidity) are close and fall into the same interval categories. The average value of the bin is then used to predict load with similar associated variables.

2.1.2. Multiple linear regression (MLR)

MLR method relates the predicted variable (baseline energy consumption) to multiple input variables. Typically, ambient temperature and time index of the monitored period are included in the input variables.

Using q_b to denote the baseline energy consumption, and U to denote the vector of input variables during the monitored period, then q_b can be expressed with the following formula:

$$q_b = q_0 + \sum_{i=1}^{s} a_i U_i \tag{1}$$

where q_0 is a constant, a is a vector of coefficients that are derived by fitting the training data with a linear curve. The notation s denotes the number of input variables, and the value of s could be different in different tools. When there is only one independent variable (degree day) in U, this method essentially becomes the degree-day method.

Building energy consumption in previous hours are often included in the input variables, to take into account the time delay caused by building thermal mass. The MLR model in Eq. (1) then becomes the autoregressive moving average with exogenous inputs (ARMAX) model in Eq. (2) below:

$$q_b = q_0 + \sum_{i=1}^{s} a_i U_i + \sum_{i=1}^{m} c_i q_{t-i} \tag{2}$$

where the notation m denotes the number of past observations. In both Eqs. (1) and (2), the coefficients q_0, a, c can be determined by so called ordinary least square (OLS) estimation techniques.

2.1.3. Support vector regression (SVR)

Support vector machine (SVM) is a data driven black box method, it gains its popularity due to its superior performance in prediction accuracy. Suppose there is an input vector X ($X_{i,j}$ denotes the jth input component in the ith sample), and a corresponding output vector Y (Y_i denotes the prediction output of the ith sample, SVM relates X with Y with the following equation:

$$Y = W \cdot \phi(X) + b \tag{3}$$

where W and b are constants, ϕ is a mapping function that maps X to a higher dimensional feature space.

SVM is special in two aspects: first, it tries to minimize not only training error, but also structure risk described by the norm of W; second, SVM ignores training samples whose training errors are smaller than a predefined constant ε. By introducing slack variables ξ_i and ζ_i^*, SVM calculates W and b with the following algorithm.

$$\left.\begin{array}{l} \underset{\zeta_i,\zeta_i^*,W,b}{\text{Minimize}} : \frac{1}{2}\|W\|^2 + c\frac{1}{N}\sum_{i=1}^{N}(\zeta_i + \zeta_i^*) \\ \text{subject to } Y_i - W\cdot\phi(x_i) - b \le \varepsilon + \zeta_i \\ W\cdot\phi(x_i) + b - Y_i \le \varepsilon + \zeta_i^* \\ i = 1,\ldots,N, \zeta_i \ge 0, \zeta_i^* \ge 0 \end{array}\right\} \tag{4}$$

2.1.4. Gaussian process regression (GPR)

The principle of GPR lies in the concept that the observations y_i ($i = 1,\ldots,n$) can be regarded as a point sampled from a multivariate distribution with n independent input vectors X_j ($j = 1,\ldots,n$). To do so, the observations are correlated with a vector of mean values m and covariance function K of X with the normal distribution function, as shown in Eq. (5).

$$P(y; m, k) = \frac{1}{(2\pi)^{n/2}|K(X,X)|^{1/2}} \times \exp\left(-\frac{1}{2}(y-m)^T K(X,X)^{-1}(y-m)\right) \tag{5}$$

where the covariance function K is calculated by Eq. (6):

$$K = \begin{vmatrix} k(x_1,x_1) & k(x_1,x_2) & \cdots & k(x_1,x_n) \\ k(x_2,x_1) & k(x_2,x_2) & & k(x_2,x_n) \\ \vdots & \vdots & \ddots & \vdots \\ k(x_n,x_1) & k(x_n,x_2) & & k(x_n,x_n) \end{vmatrix} \tag{6}$$

in which the kernel function k can be polynomial function, radial basis function, sigmoid function, etc.

To take into account the uncertainty in prediction of y, a white noise σ is assumed, thus the covariance of y becomes

$$\text{cov}(y) = K(X,X) + \sigma^2 \tag{7}$$

With above assumption, and denote the new observation variable as x^*, a new prediction of y^* can be formulated as below.

$$y^* = \sum_{i=1}^{n} \alpha_i k(x_i, x^*) \tag{8}$$

$$\alpha_i = (K(X,X) + \sigma^2 I)^{-1} y_i \tag{9}$$

2.1.5. Artificial neural network (ANN)

The field of artificial neural network is derived from neurobiological studies. There are many types of neural networks, such as feed forward network, recurrent network, and radial basis function (RBF) network. An ANN is composed of multiple layers of neurons and functions that connect them, as shown in Fig. 1. Various functions are available to connect these neurons, and linear function and sigmoid function are the most commonly used. Once an ANN together with its functions has been trained with a proper amount of data sets, it can be used to predict the output and achieves good accuracy even if the input–output relationship remains unknown.

Regarding the structure of an ANN, the feed forward network, recurrent network and RBF network are the most common

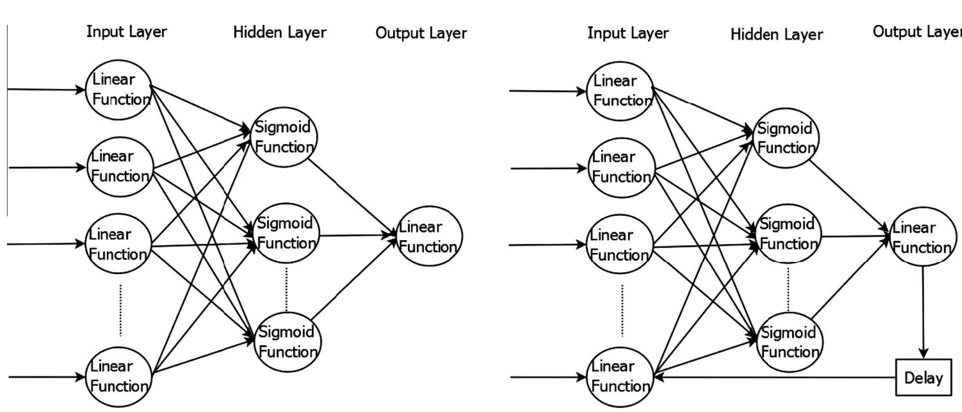

Fig. 1. A typical three layer feed forward network (left: feed forward network, right: recurrent neural network).

configurations. A typical feed forward neural network is shown in Fig. 1. In this network, the information flows from input nodes to output nodes, there is no loop or cycle in the network. On the other hand, a recurrent network allows cycles from output nodes to input nodes, thus enables learning from past experience (as shown in Fig. 1). Besides sigmoid function, radial bases function (RBF) is also widely as activation function. If RBF function is used, the neural network is also terms as RBF network.

Once a structure is determined, the weights and parameters in activation functions can be trained via a back propagation algorithm, given sufficient training data.

2.1.6. Decision tree (DT)

In the artificial intelligence (AI) area, decision tree method is traditionally used for classification. If regarding the energy benchmarking problem as a classification problem, this method can also be used to benchmark building energy consumption against its historical data. When using this method, the independent variables (weather, building type, room area, etc.) are firstly selected and converted into either categorical parameters or numerical values, further, the target variable (historical energy consumption) is also converted into categorical parameter. These historical data can serve as the training data to derive an energy consumption decision tree, based on popular decision tree algorithms. This decision tree can then be used to benchmark current energy consumption based on the new input variables.

2.2. Gray box method

2.2.1. Bayesian network (Bayesian network)

Bayesian network represents a set of random variables and their conditional dependencies with a probabilistic graphical model. Given a set of random input variables x_i ($i = 1,\ldots, N$), a Bayesian network specifies how input variable are connected with each other, and a conditional distribution for each node. A typical conditional distribution is shown in Eq. (5), where θ denotes a set of unknown model parameters, and U_i denotes the parent nodes for node x_i.

$$P(x_i, \theta) = P(x_i|U_i, \theta) \tag{10}$$

If some of the x (denoted by Z_i) are unobservable, which is often the case, the equation above becomes

$$P(x_i, \theta) = \sum_{Z_i} P(x_i, Z_i|U_i, \theta) \tag{11}$$

Given a set of training data, the unknown model parameters θ as well as the missing data Z can be estimated using expectation maximization (EM) algorithm. Typically, Eq. (6) is transformed to its log counterpart to avoid the problem of too many missing data. Denoting the missing variables as h and visible variables as v, the log form of Eq. (6) is shown in Eq. (7). By maximizing $\log P(x,\theta)$, the distribution of h and θ can be estimated.

$$\log P(x, \theta) = \log P(\theta) + \sum_i \int_{h_i} \log P(h, v|\theta) dh \tag{12}$$

2.2.2. RC network for air conditioning load

RC network method uses an analogy to the electrical circuit to model heat transfer through structures. R and C in the thermal transfer context stand for the thermal resistor and thermal capacitor respectively. The most common network configuration for the wall, roof and floor is 3R2C network, and for modelling internal mass 2R2C network is typically used (as shown in Fig. 2). The number before R and C stands for the number of resistors and capacitors, respectively. A three-step approach is used to model the cooling load. First, frequency analysis is used to estimate the parameter for 3R2C network. Second, an optimization approach combining the line search and sequential quadratic programming (SQP) is used to find the optimal parameter for 2R2C network. Finally, the estimated network structure is used to predict the cooling load of the building.

2.3. White box method

In white box method, the modeller submit a set of input parameters (typically building design parameters) to a calculation tool,

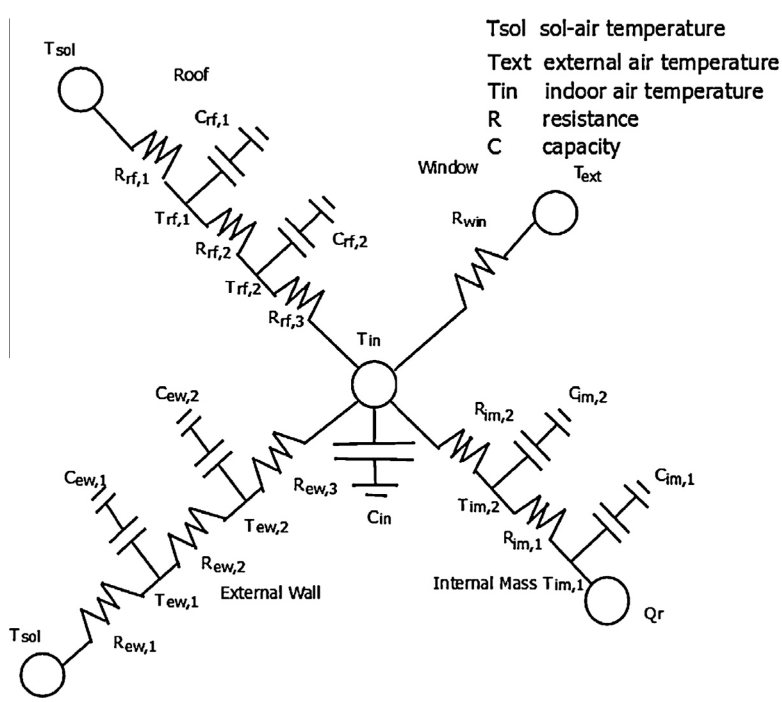

Fig. 2. RC network.

which then does the calculation and send monthly or hourly energy consumption as output. In the order of increasing complexity, there is normative method, modified bin method and equation based energy simulation method.

2.3.1. Normative method

As mentioned above, in normative method, the energy consumption baseline is calculated a simplified energy flow model, taking into account the building geometry, location, envelope material, types and configurations of major building systems (heating, cooling, humidifying, lighting, pump, fan, domestic hot water, etc.), and even the electricity generation efficiency on the primary energy side. During the calculation, the normative method has to make assumptions on key parameters during the calculation purpose. If the normative method specifies the parameter values according to input information using a standardized way, based on an empirical study of a large number of buildings, then the prediction result is the intended energy performance of the building [14]. However, if these parameters are not pre-assumed, but guessed through model calibration using monitoring data, then this method sets up a benchmark model against the past of the building [15].

2.3.2. Idealized model based method

An idealized building model based method is similar to normative method in that both methods follow a simplified and standard procedure to calculate the energy consumption benchmark, and thus require relatively few design parameter inputs. This standard calculation procedure is typically derived from first principle based method, but with some empirical assumptions to simplify the calculation. However, different from building energy standards which is established by a group of experts, idealized building model is often established by one or two researchers, and is tailored to some specific applications [16,17].

2.3.3. Modified bin method

Knebel developed a modified bin method, which is a simplified energy analysis calculation method [7]. In this method, outdoor temperature is characterized using four typical temperatures: peak cooling temperature (T_{pc}), intermediate cooling temperature (T_{ic}), intermediate heating temperature (T_{ih}), peak heating temperature (T_{ph}). For each temperature, a building load is calculated by considering the following components: solar through glass, conduction through glass, conduction through walls and roof, solar through walls and roof, lights, people, equipment and infiltration. By correlating each component linearly with outdoor temperature, the total load profile is also linearly related with outdoor temperature. Thus, total heating or cooling demand is the summed product of temperature hours and corresponding load. On the system side, modified bin method contains calculation methods for a large set of commonly used HVAC systems, such as heating/cooling coil, fan, mixing air box, chiller, boiler, cooling tower, etc.

2.3.4. Detailed energy simulation method

A detailed energy simulation method calculates the energy consumption for heating, cooling, ventilation, and lighting equipment associated with a building based on first principle models and detailed input information. Take heating and cooling load calculation as an example, it calculates the heating and cooling loads to maintain thermostat set points based on the orientation, geometry, location, envelope, climate condition and internal thermal load of the building, etc.

Crawley compared twenty major building energy simulation programs (DOE-2.1E, EnergyPlus, DeST, etc.) based on detailed simulation method, regarding their general modelling features. It is found that the comparison is difficult as there is no common language when describing the capabilities of different software. However, following capabilities are commonly found in a detailed energy simulation method: heating/cooling load calculation considering conduction, convection, long wave and short save radiation; ventilation calculation considering wind pressure coefficient based multiple zone air flow network; HVAC system performance calculation considering an inter-connection between plants, distribution systems and terminal units; tight coupling between instantaneous building load and HVAC system performance, etc. [18].

3. Application of benchmarking methods in continuous commissioning

All of the methods introduced above have been applied to benchmark building energy consumption. A review of detailed application of these methods is given in this section. A summary of these applications is given in Table 1.

3.1. Black box method

Bin method may be one of the earliest studied methods for load prediction, and has been used in two existing continuous commissioning (CC) tools: WBD and PACRAT [19,20]. It is used for baseline calculation and fault detection at the building level. While WBD is more of a research tool, PACRAT is commercially available through facility dynamics engineering [21], and has been used in many different projects [22].

MLR (or ARMAX) model has been successfully used to predict peak electricity load, hourly air conditioning load, and hourly total building energy consumption [23–26]. Jacob [25] tested the performance of linear regression methods in building fault detection applications. They find that by introducing the rate of change of indoor air temperature (ΔT) as an independent variable, the coefficient of determination (R^2) between predicted and measured energy consumption can be improved from 0.2 to 0.7. Use of a hierarchical agglomerative clustering algorithm is found to further increase the value of R^2.

SVR has also been found useful in predicting hourly cooling load and monthly utility bill [27–30]. Researchers applied SVM to predict cooling load of simulated buildings in Guangzhou city, good prediction accuracy was observed i [27–29]. Dong applied SVR to predict monthly utility bill for four commercial buildings in Singapore based on weather data (ambient temperature, relative humidity and global solar radiation), prediction accuracy within 4% was achieved [30].

GPR is a unique regression technique in that it explicitly calculates the uncertainty of the prediction result, thus is more useful when the uncertainty of input parameters is large. Since building energy modeller often lacks confidence in choosing a proper value for certain parameters, such as the envelope insulation level (U value), and the air leakage rate, an explicit quantification of the impact of input parameter uncertainty on the outcome could be attractive. For this reason, it has been proposed by several researchers as a meta-modelling technique for building energy consumption estimation [31,32]. Heo concluded that GPR can capture the complex behavior of energy consumption better than MLR, and are more suited to handle the problem of uncertainty [31]. Manfren compared the performance of MLR, GPR and a detailed energy simulation model, and found that the performance of GPR is much closer to that of the detailed simulation model than MLR method. Thus, it is concluded that GPR is an effective technique in cases where detailed simulation model is difficult to establish [32].

ANN has been tested by numerous researchers for predicting cooling load [24,28], hot-water heating load [33], space heating

Table 1
Summary of energy benchmarking models.

Method	Input	Time resolution level	Application	Tool
Bin method	Day of the week, hour of the day, weather data	Multiple resolution	Fault detection [19,20]	WBD, PACRAT
Linear regression	Day type, weather data, historical data	Multiple resolution	Fault detection [25], load prediction [23–24,26]	/
Support vector regression	Weather data	Multiple resolution	Prediction of cooling load [27–29], monthly utility bill split [30]	/
Gaussian process regression	Weather data, other selected explanatory variables	Multiple resolution	Retrofit analysis [31], replacement of detailed simulation [32]	/
Artificial neural network	Weather data, time, DHW/heating system and equipment properties, energy consumption patterns, dwelling characteristics	Hourly	Prediction of cooling load [24,28], hot water heating load [33], space heating load [33,34], total energy consumption [35], fault detection [36]	/
Decision tree	Weather, building type, ownership of electric appliance, building area	Annually	Electricity prediction [37], total energy prediction [38]	/
Bayesian network	Weather data, known parameters, prior distributions of unknown parameters, historical energy consumption data	Daily	Energy consumption estimation [39]	/
RC network	Weather data, historical energy consumption data	Hourly	Building heating/cooling load predication [40,41], demand control [42]	/
Normative	Weather data, simplified building design parameters	Hourly, monthly	Energy source planning, energy policy analysis [43], retrofit analysis [15]	EPSCT
Idealized model based	Weather data, simplified building design parameters	Hourly	Energy consumption benchmark [16,17], optimal control [46]	/
Modified bin method	Weather data, simplified building design parameter	Hourly	Fault detection [47,48]	ABCAT
Detailed simulation	Weather data, detailed building design parameters	Sub-hourly	Monthly utility bill split [52], Fault detection [49], retrofit analysis [50–51,54], load prediction [53]	EnergyPlus Esp-r, DOE-2.1E, DEST, etc.

[33,34], and total energy consumption [35]. The results show that ANN has sufficiency accuracy for hot-water heating, cooling and space heating load prediction purposes. Based on its prediction capability, ANN has been deployed to detect faults at the building level [36].Decision tree (DT) method is a relatively new method for predicting building energy use. Tso and Yau [37] applied this method to predict electricity usage of different buildings in Hong Kong, and found that DT method has better performance than neural network method and regression analysis method. Bon the residential building energy consumption data collected in Japan, Yu et al. [38] derived a decision tree which can predict whether the annual energy consumption will be high (between 441.5 MJ/m^2 yr and 707 MJ/m^2 yr) or low (below 441.5 MJ/m^2 yr) based on multiple input information.

Holcomb compared three black box methods (MLR, SVR and ANN) in load prediction performance using simulation data, and he found that while MLR and SVR showed similar prediction accuracy, ANN had the worst performance among the three methods [9]. Besides Holcomb, other researchers also found the superior performance of SVR over ordinary back propagation ANNs [28]. It should be noted that the accuracy of ANN can be significantly improved by deploying statistical procedures (hypothesis testing, information criteria, cross validation, etc.) and optimizing the input parameters, number of hidden layers and hidden units [35].

3.2. Gray box method

Bayesian network has been applied to estimate building energy consumption for three retail and food service buildings in Florida's Walt Disney World Resort [39]. In this model, some of the known parameters (building geometry, operation schedule, envelope material, etc.) are given a prior distribution, while other parameters (air exchange per hour, window solar heat gain coefficient, latent heat removal efficiency, etc.) are estimated using the EM algorithm. It is concluded that Bayesian network is more applicable for benchmarking building energy consumption on a large scale, and especially suitable for cases where part of the parameters are unknown or some of the monitoring data is missing.

RC network is mainly used to calculate air conditioning load. It has been used by multiple researchers for load prediction and demand control purposes [40–42]. Since this method requires less training data (one or two week traiing data is sufficient), and less input parameters (mainly weather data and internal temperature set point), it is an effective and robust method for air conditioning load estimation. Li and Huang found that RC network showed better adaptability to change of temperature set points, compared with black box methods [24].

3.3. White box method

Normative method has been adopted and applied by some research groups. The building technology group in Georgia Institute of Technology developed a normative energy benchmarking tool named Energy Performance Standard Calculation Toolkit (EPSCT), to benchmark both new designed building and existing buildings [43]. In this tool, they used different standards to calculate each sub-item in total building energy consumption. For example, they use EN ISO 13789 for transmission and ventilation heat transfer, EN 15241 and EN 15242 for ventilation for buildings, EN 15243 for cooling and ventilation systems, etc. [44,45]. Lee applied this tool in two areas: energy source planning of a campus, and energy policy analysis, and concluded that this tool is capable for these two purposes [43]. Heo (2012) calibrated a normative model using real measurement data from a building located in Cambridge, UK. She concluded that a calibrated probabilistic normative model is as effective in retrofit analysis decision making as legacy software, such as EnergyPlus [15].

Although maybe not explicitly termed as such, idealized model based method has been used by several researchers for energy benchmarking purposes [16,17,46]. Federspiel proposed a simplified building energy benchmark model for laboratory buildings. In this model, it is assumed that the building has perfect insulation (no heat conduction or convection with external environment), and allows no solar radiation into the building. Thus, the total cooling and heating load is mainly composed of ventilation load and internal heating/cooling load. With this method, user is required to define just a few input parameters: plan area of lab space and non-lab space, fraction of lab-space that is air conditioned, location, etc. [16]. Mui proposed a simple energy benchmarking method for ventilation system in air conditioned offices, which relates the required ventilated rate with the CO_2 concentration in the office, and calculates the energy consumption accordingly [17]. Lv proposed a new model based method, which can take into account multiple uncertainty sources and achieve high accuracy in modelling the transient building thermal behavior [46].

Modified bin method is one of the few white box methods that have been applied for fault detection and diagnostics purposes in real buildings. Researchers in Texas A&M University developed a tool named Automated Building Commissioning Analysis Tool (ABCAT) for fault detection at whole building level [47]. In addition to the original modified bin method, ABCAT adds some other characteristics: calculation of electric humidification load, calculation of leakage flow rates, simplified thermal mass considerations. For calibration purpose, ABCAT introduces a bias item to reduce the difference between measured and predicted energy consumption, which is estimated by minimizing mean bias errors (MBE). ABCAT has been tested in buildings in different locations (TX, NY, NE) and proved to be successful in detecting energy consumption abnormalities [48].

Detailed energy simulation method is probably by far the most studied and widely applied simulation method. It has been used for fault detection, retrofit analysis, utility bill splitting, etc. [49–54]. However, calibrating a detailed energy simulation program is a difficult matter, for two reasons: first, there are too many uncertain inputs and parameters that can affect the results; second, available data for calibration is typically little (often only annual utility data is available). To solve the problem, a sensitivity analysis technique – Monte Carlo (MC) is often deployed [55]. ASHRARE research project RP-1051 proposed a five step calibration methodology: preliminary input file preparation, parameter space reduction by walk-through audits, Monte-Carlo based coarse grid calibration, solution refinement through guided search, and finally analysis using multiple plausible solutions [56]. Another approach proposed by Raftery focuses on calibrating the detailed model iteratively and on a continuous basis [57].

4. Discussion

4.1. How to choose a proper benchmarking method

A comparison between different energy benchmarking methods has been made regarding: input data requirement, modeller experience requirement, calibration effort requirement, and training data requirement. Each characteristic is grouped into three levels: low, medium and high. A comparison between different benchmarking methods on these requirements has been made in Table 2. In general, as the model goes from black box category to white box category, the information input requirement increases, the requirement for calibration also increases, although the requirement on training decreases. On the other hand, the learning curves of these methods are less dependent on the method category, and are more related with the complexity of the model structure. While linear regression, bin method (BM) and decision tree (DT) method are conceptually simple and thus easy to use, black box methods are challenging due to the embedded statistical knowledge and concepts, and white box methods require certain extent of knowledge in the discipline of building physics. Detailed energy simulation method is arguably the most challenging method due to its complex model structure and solution procedure.

Furthermore, a flow chart (Fig. 3) has been made to help reader select a proper benchmarking method. When detailed information of the building is available, preference should be given to white box methods (detailed energy simulation method, normative method, modified bin method or idealized model based method), since the principles of these methods are easier to understand and the results are thus easier to interpret. Among white box methods, detailed energy simulation model usually takes much longer development time, and also requires larger computational resources, thus it should be only used when there is an extremely high requirement on understanding the temporal and spatial characteristics of the energy consumption data. In case detailed information is not available, then either gray box or black box methods should be used, choosing which method depends on the requirement on the energy benchmarking target, level of dynamic characteristic, confidence in using these methods, ability to quantify uncertainty, etc. For example, while RC network method should be chosen to control indoor temperature setpoint, since it has a better accuracy in predicting the dynamic characteristics of the heat exchange inside the building; black box methods are

Table 2
Comparison between different benchmarking methods.

Method	Quantity of input data requirement	Modeller experience requirement	Calibration effort requirement	Quantity of training data requirement
Linear regression	Low	Low	Low	High
Support vector regression	Low	Medium (familiar with statistical concepts)	Low	High
Gaussian process regression	Low	Medium (familiar with statistical concepts)	Low	High
Artificial neural network	Low	Medium (familiar with data processing)	Low	High
Bin method	Low	Low	Low	High
Decision tree	Low	Low	Low	High
Bayesian network	Medium	Medium (familiar with probability theory)	Medium (few parameters)	Medium
RC network	Low	Medium (familiar with thermal dynamics theory)	Medium (few parameters)	Medium
Normative method	Medium	Medium (familiar with building physics)	Medium (relatively more parameters)	Medium (can be Low, if accuracy is not required)
Idealized model based	Medium	Medium (familiar with building physics)	Medium (relatively more parameters)	Medium
Modified bin method	Medium	Medium (familiar with building physics)	Medium (relatively more parameters)	Medium
Detailed energy simulation	High	High (familiar with building physics and the particular software)	High (most parameters)	Low

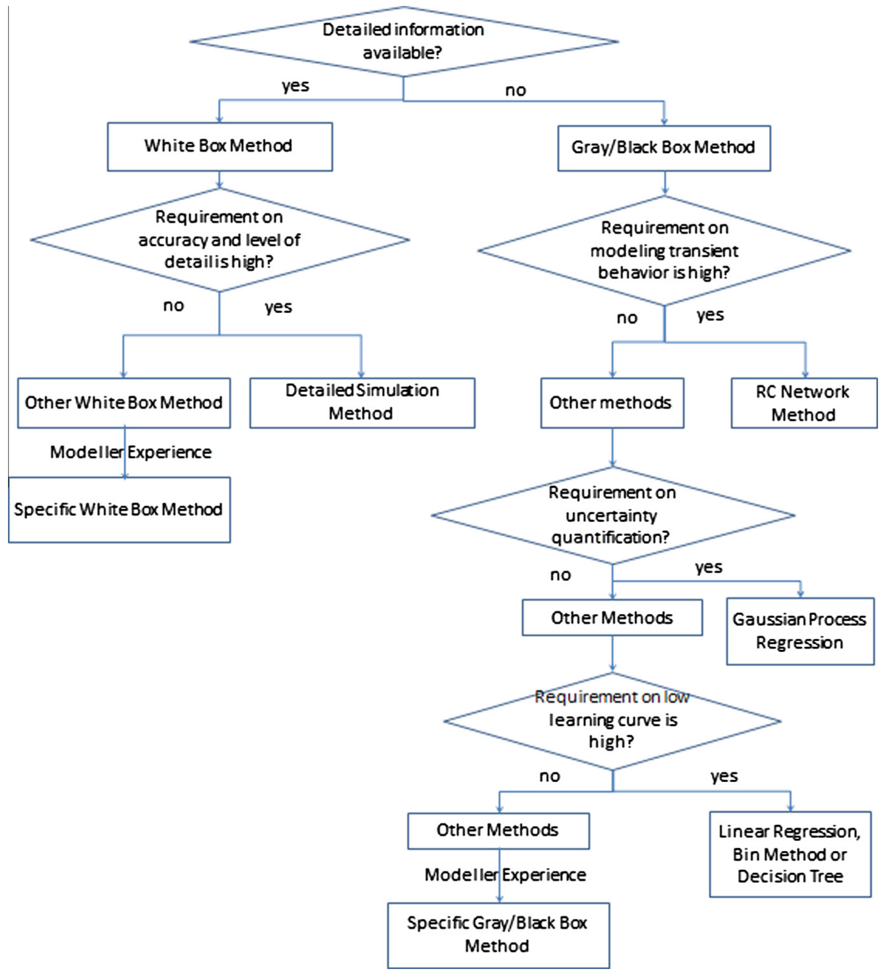

Fig. 3. Choose a proper energy benchmark method.

better suited when the total building energy consumption needs to be benchmarked Since the achievable benchmarking accuracy depends a lot on the experience of the modeller, the modeller himself/herself should be another key factor in choosing which method to use (see Fig. 3).

4.2. Performance of black box methods

Due to its convenience and quick modelling, black box methods are good alternatives to detailed energy simulation method. Although all derived from data mining techniques, the principle embedded in different black box methods still causes a difference in their characteristics. Bin method is the simplest method, yet perhaps one of the most widely applied methods. The tools (WBD and PACRAT) that deploy it have been applied in numerous buildings for continuous commissioning. Multiple linear regression (MLR) technique is the simplest technique, and has been adopted by ASHRAE as a standard Measurement & Verification (M&V) technique [58]. Artificial neural network (ANN) method is arguably the most widely used non-linear regression method in building continuous commissioning, and has achieved success in many applications. However, as it requires tweaking the inputs, network structure and weight parameters, its accuracy can't be guaranteed. In some cases, it could perform worse than MLR. Support vector regression (SVR) is a unique regression method in that it optimizes both the model structure and the estimator error, and its performance has been proved by many researchers, including winning entry in load prediction competitions, thus is worth investigating in real applications. Gaussian process regression (GPR) is the only method among black box methods that explicitly calculates the uncertainty of the estimation result. Combining this with its nonlinear regression nature and ability to capture complex behavior, it should get wider application in cases requiring risk analysis.

4.3. Performance of gray box methods

While RC network has its root in building physics, and has been used by relatively more researchers for building heating/cooling load predictions; Bayesian network method is derived from statistical theory and has only been popular in recent years. It is obvious that Bayesian network has a broader use than RC network method, since the latter is only applicable in heating/cooling load calculation. Furthermore, Bayesian network method has a strong capability in handling measurement data errors. However, it should be noted that the structure of a Bayesian network can only be applied in building with similar building use pattern and energy consumption profiles, therefore is less general than RC network method.

4.4. Performance of white box methods

Detailed simulation method is probably the most widely used method for energy estimation in design stage. Due to its comprehensiveness and wide acceptability, it is often used as a comparison case when testing new benchmarking methods. Researchers have shown that simplified methods can perform as effective as detailed simulation method in many cases for energy benchmarking purpose [31,32]. Developers of ABCAT have shown that modified bin method performs satisfactorily for energy benchmarking purposes in testing cases. Normative method is another white box method with potential for use in continuous commissioning, since it is as effective as a detailed simulation method in retrofit analysis [15].

5. Conclusion

In this paper, a total of twelve methods for benchmarking building energy consumption are reviewed, including six black box methods, two gray box methods and four white box methods. It is found that many methods, although simple, can achieve satisfactory performance. Choosing a proper method should be based on project requirements, available inputs, available monitoring data, and modeller experience. While white box methods should be given higher priority due to the transparency in the calculation procedure, black or gray box methods may be sufficient if enough training data is available. A flow chart has been proposed to help modeller choose a proper benchmarking method.

Acknowledgement

The work described in this paper was fully supported by a Grant from the Shanghai Committee of Science and Technology, China (Project No. 12230708900).

References

[1] IEA, Key world energy statistics; 2009.
[2] Philp Haves. Fault modelling in component-based HVAC simulation. In: Fifth international IBPSA conference, Prague, Czech, September 8–10, 1997.
[3] Xiao F, Wang S. Progress and methodologies of lifecycle commissioning of HVAC systems to enhance building sustainability. Renew Sustain Energy Rev 2009;13:1144–9.
[4] Claridge DE, Culp CH, Liu M, et al. Campus-wide continuous commissioning of university buildings, ACEEE 2000 Summer Study on Energy Efficiency in Buildings, Pacific Grove, U.S., August 20–25, 2000.
[5] Federal Energy Management Program (FEMP). Continuous Commissioning Guidebook for Federal Managers. <http://eber.ed.ornl.gov/CommercialProducts/>; 2013.
[6] Djuric N, Novakovic V. Review of possibilities and necessities for building lifetime commissioning. Renew Sustain Energy Rev 2009;13:486–92.
[7] Knebel DE. Simplified energy analysis using the modified bin method – prepared for the American Society of Heating, Refrigerating and Air-Conditioning Engineers, ASHRAE, Atlanta, GA; 1983.
[8] Crawley DB, Lawrie LK, Pedersen CO, et al. Energy plus: energy simulation program. ASHRAE J. 2000;42:49–56.
[9] Holcomb 9 D, Li W, Seshia SA. Algorithms for green buildings: learning-based techniques for energy prediction and fault diagnosis, UCB/EECS-2009-138. University of California at Berkeley; 2009.
[10] Al-Homoud MS. Computer-aided building energy analysis techniques. Build Environ 2001;36:421–33.
[11] Roth K, Llana P, Westphalen D, et al. Automated whole building diagnostics. ASHRAE J 2005;47:82–4.
[12] Chung W. Review of building energy-use performance benchmarking methodologies. Appl Energy 2011;88:1470–9.
[13] Zhao H, Magoules F. A review on the prediction of building energy consumption. Renew Sustain Energy Rev 2012;16:3586–92.
[14] Augenbroe G, Park C. Quantification methods of technical building performance. Build Res Inf 2005;33:159–72.
[15] Heo Y, Choudhary R, Augenbroe GA. Calibration of building energy models for retrofit analysis under uncertainty. Energy Build 2012;47:550–60.
[16] Federspiel C, Zhang Q, Arens E. Model-based benchmarking with application to laboratory buildings. Energy Build 2002;34:203–14.
[17] Mui KW, Wong LT, Law LY. An energy benchmarking model for ventilation systems of air-conditioned offices in subtropical climates. Appl Energy 2007;84:89–98.
[18] Crawley DB, Hand JW, Kummert M, et al. Contrasting the capabilities of building energy performance simulation programs. Build Environ 2008;43:661–73.
[19] Katipamula S, Pratt R, Chassin D, et al. Automated fault detection and diagnostics for outdoor-air ventilation systems and economizers: methodology and results from field testing. ASHRAE Trans 1999;105:555–67.
[20] Santos J, Brightbill L. Automated diagnostics from DDC data – PACRAT. In: Proceedings of the 8th National conference on building commissioning, Kansas City, U.S., May 3–5, 2000.
[21] Facility dynamics. <http://www.facilitydynamics.com/software.html> ; 2013.
[22] Friedman H, Piette MA. Comparative guide to emerging diagnostic tools for large commercial HVAC systems, Lawrence Berkeley National Laboratory Report (No. 48629); 2001.
[23] Papalexopoulos AD, Hesterberg TC. A regression-based approach to short-term system load forecasting. IEEE Trans Power Syst 1990;5:1535–47.
[24] Li Z, Huang G. Re-evaluation of building cooling load prediction models for use in humid subtropical area. Energy Build 2013;62:442–9.
[25] Jacob D, Dietz S, Komhard S, et al. Black-box models for fault detection and performance monitoring of buildings. J Build Perform Simul 2010;3:53–62.
[26] Yiu JC, Wang S. Multiple ARMAX modelling scheme for forecasting air conditioning system performance. Energy Convers Manage 2007;48:2276–85.
[27] Li Q, Meng Q, Cai J, et al. Applying support vector machine to predict hourly cooling load in the building. Appl Energy 2009;86:2249–56.
[28] Li Q, Meng Q, Cai J, et al. Predicting hourly cooling load in the building: a comparison of support vector machine and different artificial neural networks. Energy Convers Manage 2009;50:90–6.
[29] Li X, Lu J, Ding L et al. Building cooling load forecasting model based on LS-SVM. In: Asia-Pacific conference on information processing, Shenzhen, China, July 18–19, 2009.
[30] Dong B, Cao C, Lee SE. Applying support vector machines to predict building energy consumption in tropical region. Energy Build 2005;37:545–53.
[31] Heo Y, Zavala VM. Gaussian process modelling for measurement and verification of building energy savings. Energy Build 2012;53:7–18.
[32] Manfren M, Aste N, Moshksar R. Calibration and uncertainty analysis for computer models – a meta-model based approach for integrated building energy simulation. Appl Energy 2013;103:627–41.
[33] Aydinalp M, Ugursal VI, Fung AS. Modeling of the space and domestic hot-water heating energy-consumption in the residential sector using neural networks. Appl Energy 2004;79:159–78.
[34] Mihalakakou G, Santamouris M, Tsangrassoulis A. On the energy consumption in residential buildings. Energy Build 2002;34:727–36.
[35] Karatasou S, Santamouris M, Geros V. Modeling and predicting building's energy use with artificial neural networks: methods and results. Energy Build 2006;38:949–58.
[36] Magoules F, Zhao H, Elizondo D. Development of an RDP neural network for building energy consumption fault detection and diagnosis. Energy Build 2013;62:133–8.
[37] Tso GKF, Yau KKW. Predicting electricity energy consumption: a comparison of regression analysis, decision tree and neural network. Energy 2007;32:1761–8.
[38] Yu Z, Haghighat F, Fung BCM, et al. A decision tree method for building energy demand modelling. Energy Build 2010;42:1637–46.
[39] Tarlow D, Peterman A, Schwegler BR, et al. Automatically calibrating a probabilistic graphical model of building energy consumption. In: 11th IBPSA Conference, Glasgow, Scotland, July 27–30, 2009.
[40] Wang S, Xu X. Simplified building model for transient thermal performances estimation using GA-based parameter identification. Int J Therm Sci 2006;45:419–32.
[41] Braun JE, Chaturvedi N. An inverse gray-box model for transient building load prediction. HVAC&R Res 2002;8:73–99.
[42] Lee K, Braun JE. Development of methods for determining demand-limiting setpoint trajectories in buildings using short-term measurements. Build Environ 2008;43:1755–68.
[43] Lee SH, Zhao F, Augenbroe G. The use of normative energy calculation beyond building performance rating systems. In: 12th IBPSA Conference, Sydney, Australia, November 14–16, 2011.
[44] ISO, ISO 13789:2007, Thermal performance of buildings – Transmission and ventilation heat transfer coefficients – calculation method; 2007.
[45] CEN, EN 15241:2007, Ventilation for buildings. Calculation methods for energy losses due to ventilation and infiltration in buildings; 2007.
[46] Lv X, Lu T, Kibert CJ. A novel dynamic modelling approach for predicting building energy performance. Appl Energy 2014;114:91–103.
[47] Curtin JM. The development and testing of an automated building commissioning analysis tool (ABCAT), Master thesis, Texas A&M University, College Station, U.S.; 2007.
[48] Bynum JD, Claridge DE, Curtin JM. Development and testing of an Automated Building Commissioning Analysis Tool (ABCAT). Energy Build 2012;55:607–17.
[49] O'Neill Z, Shashanka M, Pang X et al. Real time model-based energy diagnostics in buildings. In: 12th IBPSA Conference, Sydney, Australia, November 14–16, 2011.
[50] Torrens JI, Keane M, Costa A, et al. Multi-criteria optimisation using past, real time and predictive performance benchmarks. Simul Model Pract Theory 2011;19:1258–65.
[51] Chidiac SE, Catania EJC, Morofsky E, et al. A screening methodology for implementing cost effective energy retrofit measures in Canadian office buildings. Energy Build 2011;43:614–20.
[52] Fumo N, Mago P, Luck R. Methodology to estimate building energy consumption using EnergyPlus benchmark models. Energy Build 2010;42:2331–7.
[53] Lin H, Hong T. On variations of space-heating energy use in office buildings. Appl Energy 2013;111:515–28.

[54] Wang L, Greenberg S, Flegel J, et al. Monitoring-based HVAC commissioning of an existing office building for energy efficiency. Appl Energy 2013;102:1382–90.
[55] Saltelli A, Chan K, Scott EM. Sensitivity analysis. John Wiley & Sons Ltd.; 2000.
[56] Reddy TA, Maor I, Panjapornpon C. Calibrating detailed building energy simulation programs with measured data – part I: general methodology (RP-1051). HVAC&R Res 2007;13:221–41.
[57] Raftery P, Keane M, O'Nonnell J. Calibrating whole building energy models: an evidence-based methodology. Energy Build 2011;43:2356–64.
[58] ASHRAE, ASHRAE Guideline 14: measurement of energy and demand savings, American Society of Heating, Refrigerating, and Air-conditioning Engineers Inc., Atlanta, GA, 2002.