https://doi.org/10.1016/j.apenergy.2018.11.077

A hybrid building thermal modeling approach for predicting temperatures in typical, detached, two-story houses

Borui Cui [a], Cheng Fan [b,*], Jeffrey Munk [a], Ning Mao [c], Fu Xiao [d], Jin Dong [a], Teja Kuruganti [a]

[a] Oak Ridge National Laboratory, One Bethel Valley Road, Oak Ridge, TN 37831, USA

[b] Department of Construction Management and Real Estate, Shenzhen University, Shenzhen, China

[c] Department of Gas Engineering, College of Pipeline and Civil Engineering, China University of Petroleum (East China), Qingdao, China

[d] Department of Building Services Engineering, The Hong Kong Polytechnic University, Kowloon, Hong Kong

* Corresponding author Tel: +86-(0755)26916426.

Email address: fancheng@szu.edu.cn

Abstract

Within the residential building sector, the air-conditioning (AC) load is the main target for peak load shifting and reduction since it is the largest contributor to peak demand. By leveraging its power flexibility, residential AC is a good candidate to provide building demand response and peak load shifting. For realization of accurate and reliable control of AC loads, a building thermal model, which characterizes the properties of a building's envelope and its thermal mass, is an essential component for accurate indoor temperature or cooling/heating demand prediction. Building thermal models include two types: "Forward" and "Data-Driven". Due to time-saving and cost-effective characteristics, different data-driven models have been developed in a number of research studies. However, few developed models can predict temperatures in respective zones of a multiple-zone building with an open air path between zones e.g., an open stairwell connecting two floors of a home. In this research, a novel hybrid modeling approach is proposed to predict the average indoor air temperatures of both the upstairs and downstairs. A developed RC model is used to predict the building mean temperature, and supervised machine learning techniques are used to predict the temperature difference between the downstairs and upstairs. Compared with the measured data from a real house, the results obtained have acceptable/satisfactory accuracy. The method proposed in this study integrates the advantages of black-box and gray-box modeling. It can be used as a reliable alternative to predict the average temperatures in respective floors of typical detached two-story houses.

Keywords: Building demand management; data-driven model; supervised machine learning; particle swarm optimization;

1. Introduction

1.1 Background of research

The residential building sector accounts for approximately 37% of total U.S. electricity consumption [1]. Meanwhile, as of 2011, there were over 132 million housing units in the U.S., and of those 87% had heating, ventilation and air conditioning (HVAC) systems. In the southern United States, the percentage of homes with air conditioning approaches 100% [2]. These air conditioning (AC) units tend to exacerbate peak demand issues. For example, during the 2011 summer peak in the ERCOT (Electric Reliability Council of Texas) grid, over 50% of the total electrical load was from homes [3], whose loads were primarily driven by their air conditioning systems.

Building demand management, including both peak load management (PLM) and demand response (DR), has been proposed for building peak power and energy reduction as well as reliable operation of the electric grid [4]. Within the building sector, HVAC systems have the best potential to provide both PLM and DR [5]. HVAC peak load shifting and DR has been the focus of recent research [6, 7]. Different facilities integrated with residential HVAC systems have been applied to provide peak load shifting and DR, such as PV [8, 9, 10], battery storage system [11], active thermal storage system [12, 13], and phase change materials (PCM) [14, 15].

The performance of the above control techniques relies heavily on the quality of building model, which characterizes the properties of the building's envelope and thermal mass. Buildings contain significant thermal mass that can be utilized as a valuable distributed storage asset for a variety of applications [16]. Building thermal models can be used for different purposes, e.g., building thermal load prediction under certain set-points [17], decision of proper amount of energy to be stored in cool storage system during off-peak hours [18], precooling of building thermal mass [19], peak demand limiting [20], forecast of energy consumption in building [21], quantification of building demand flexibility [22] and fast DR control of HVAC system [23].

There are two main categories of building modeling methods according to ASHRAE: "Forward" and "Data-Driven" [24]. The most important step in selecting an energy modeling method is matching method capabilities with project requirements. A forward (white-box) model has parameters of physical significance and requires a relatively larger amount of building knowledge for practical implementations. The fundamental principles of forward models are normally based on the theory of heat transfer and thermodynamics. There are different popular building simulation programs based on this approach including EnergyPlus, Trnsys and Modelica [25]. Forward models are not practical for large-scale optimal control (virtual storage control/DR control) of building and HVAC system due to: 1. the huge effort required to manually calibrates the model parameters for buildings with limited sensor information [26]. 2. the considerable amount of building information needed for parameter setting also restricts the applicability of such kind of models. It is usually too time consuming and not cost-effective to collect such information [18].

Data-driven models adopt an inverse approach for model development, assuming that there are certain mathematical relationships between model inputs and outputs [27]. Data-driven models can be further categorized into “black-box” and “gray-box” models. A black-box model is purely derived from measured data, and model inputs and outputs may not necessarily present clear physical meaning. By contrast, gray-box models are built based on a combination of physical principles and measured data. More specifically, domain expertise is adopted to specify the model architecture and define model inputs. Measured data are then used for parameter tuning [28]. The ease of using prior knowledge of a building system within the model identification process is one obvious attraction [29].

The following sub-section introduces the state-of-art research of development of data-driven models and implication for optimal control of HVAC system and building demand management.

1.2 Literature review

One of the most popular gray-box methods is the Resistance-Capacitance(RC) model, also called lumped capacitance or network model, which is composed of electrical analogue resistances (*R*) and capacitances (*C*) [30]. Capacitance represents thermal capacity (i.e., property of objects to describe its capability to store heat), while resistance represents thermal resistance (i.e., property of objects or materials to resist the heat flow through it) [30]. The order of such kind of model is determined by the number of lumped capacitances (nodes) [31]. The physical interpretation of the parameters is dependent on how the building is divided into entities in the model [28]. The values of *R* and *C* are estimated based on samples of inputs and outputs by applying an identification algorithm, e.g., nonlinear regression algorithm, which typically minimizes a norm of either simulation errors or prediction errors [29]. The boundaries on the parameters in the identification process are normally estimated from a rough description of the building geometry and materials [26]. One obvious advantage is that the RC modeling methods allow model training using short-periods of data, e.g., one to four weeks, recorded by a limited number of sensors.

A simplified RC model for predicting single-zone building thermal load based on heat transfer of building envelope and internal mass was developed in [32]. The parameters of building thermal network models for building envelope (6R4C) were determined by frequency characteristic analysis and the parameters of thermal network models (2R2C) for lumped internal mass are identified with genetic algorithm (GA). A search for the most suitable building RC model was realized by statistical tests, i.e. likelihood ratio test and forward selection method [28]. The predicted value was compared with the “measured” indoor temperature, which was formed as the first principal component of the measurements from the indoor temperature sensors. A hierarchy of models has been formulated based on prior physical knowledge with increasing complexity, which range from 2 states/order (2R2C) to 5 states/order (5R5C), to simulate a single thermal zone for Model Predictive Control (MPC) control [33]. A method was proposed to reduce the large-state space dimension of a RC model for a multi-zone building [35]. The convection among different zones was not considered since there is no open area among the zones. A prototype energy signal tool was developed for whole-building energy use evaluation [36]. The energy consumption was modeled by a 4R1C single-zone model and the uncertainty in energy

consumption was quantified by a Monte Carlo exploration of the influence of model parameters on energy consumption. A method was proposed to improve the accuracy of single-zone RC models by considering unmeasured disturbances [29]. Both physical parameters in a RC model and disturbance model which characterizes the unmeasured inputs to the RC model were estimated. A semi-dynamic building envelope model is developed in which a RC model (5R2C) is the first part to predict heat transfer along the width of the active pipe-embedded envelope [39]. The results were compared with the output of a CFD model which is used as a virtual experiment for validating the semi-dynamic model. A new singe-zone wall system with structure of three-layer sandwich-type panel with outer layers consisting shape-stabilized phase change material (SSPCM) wallboard and conventional brick was developed [40]. The simplified dynamic model represented the brick layer by three resistances and two capacitances (3R2C) and the SSPCM layer by four resistances and two capacitances (4R2C). A 2R2C model was developed to calculate average cooling demand reduction in different zones of a commercial building simulated by TRNSYS when the developed fast DR strategy using both active and passive building thermal storages was applied [41]. It helped to predict the effective storage capacity and storage efficiency of the building mass. The above developed model was also taken as a part of the identification process of active cool thermal energy storage capacity and average power reduction for building demand management [42].

Recently, black-box models have gained increasing interests due to their capability in analyzing large-scale data and flexibility in practical applications. In the building field, black-box models have been primarily used to predict whole-building energy consumptions, indices for system performance, and indoor environment [43, 44, 45]. The techniques used can be generally categorized into two types, i.e., statistical methods and supervised machine learning. Commonly used statistical methods include multiple linear regression and time series modeling. As an example, the most straightforward approach to building cooling load prediction is to develop a linear regression model between building cooling load and indoor-outdoor temperature differences [46, 47]. Considering that building operations generally present lagging effects, time series modeling methods, such as autoregressive model with extra input (ARX) and autoregressive integrated moving average (ARIMA) models, have been used to take into account the effect in previous time steps [48, 49, 50, 51]. While statistical methods are relatively easy to implement, they can only capture linear relationships among building variables. Since building operations are typically complicated and nonlinear, the resulting accuracy can be poor. Supervised machine learning algorithms have been used as solutions [52]. The most widely used techniques include support vector machines (SVMs) [53, 54], artificial neural networks (ANNs) [55, 56], decision trees [57] and ensemble learning [58]. As an example, seven different machine learning techniques, including Feed-Forward Neural Network (FFNN), Hierarchical Mixture of Experts (MHE) and etc., were tested and compared based on the data from a residential house [59]. The Least Squares Support Vector Machines (LS-SVM) resulted in the best prediction performance in terms of 1-hour ahead building energy consumption. Different machine learning algorithms, such as ANN and SVM, were applied to predict the profiles of non-AC loads, which were then used as inputs to the RC model and AC regression model [26]. The research results validated the usefulness of machine learning in modeling complicated and nonlinear

relationships. In most above research works; the prediction targets were the thermal dynamics and power performance in single-zone buildings.

1.3 Research gaps and research objective

Previous studies mainly focused on the prediction of average temperature or cooling demand in single-zone buildings, e.g. one node to represent the average indoor temperature [29, 32, 33, 36, 40, 48, 54, 56], as well as prediction of overall mean temperature (average indoor air temperature of multiple floors)/overall cooling demand for multiple-zone buildings [28, 41, 42, 57, 58]. Few developed models can predict mean temperatures in respective zones of in-situ multiple-zone building with an open air path between zones, e.g. the detached two-story house with open stairwell. Many existing detached two-story houses in the U.S. use one AC system combined with supply duct dampers to control the respective temperatures in downstairs and upstairs. Alternatively, new homes are often equipped with two separate AC systems, i.e., one for upstairs and one for the downstairs. Under both situations, the ability to predict respective temperatures of both the upstairs and downstairs is required to ensure the temperatures stay within the respective (potentially different) comfort limits specified by the homeowner during AC system DR or optimal control in common detached two-floor houses.

However, it could be very difficult to specify a comprehensive gray-box model to take into the air mass movement and the corresponding heat transfer in the stair area due to the convection effect.

Meanwhile, there are few studies on the development of black-box building models for residential buildings, rather than commercial buildings [53, 59]. Most developed models were used to estimate whole building monthly electrical consumption based on low granularity data sets. Therefore, there is a need to investigate the usefulness of state-of-the-art machine learning techniques for predicting the dynamic performance in residential buildings based on high granularity data input, which is essential for optimal control or fast demand response control.

In this research, a creative hybrid modeling approach, using both gray-box and black-box modelling methods, is proposed to realize prediction of average indoor air temperatures of both the upstairs and downstairs of a typical detached two-story house. The developed RC model is used to predict the overall mean indoor air temperature (average indoor air temperature of both floors) and the data-driven model is used to predict the temperature difference between the downstairs and upstairs. The separate temperatures in the downstairs and upstairs are achieved accordingly.

Several major innovative aspects are involved in this study:

1. First, this research provides a practicable and effective solution for prediction and control of the indoor air temperatures of downstairs and upstairs in two-story houses with open stairwells.
2. Second, this research shows that the adopted black-box methods can efficiently and accurately predict the dynamic building thermal performance while, in most previous research, black-box models were used to estimate monthly electrical consumption estimation based on low-granularity data in residential homes.

3. Third, the developed hybrid building thermal modeling approach can be easily applied in different scenarios, such as HVAC system operation monitoring, retrofit projects, and demand flexibility identification.
4. Fourth, the performance of different black-box modeling methods is compared in detail based on the data from a real typical detached two-story house.

The following sections in this paper begin with a description of the typical two-story house in the U.S. and the installed AC system. The natural convection phenomenon and developed hybrid modelling approach is then introduced in Section 3. In Section 4, the outline of gray-box model is described. The testing results, i.e. overall mean indoor temperature, and the corresponding accuracy are shown. It is followed by the development of black-box models in Section 5. The performance of different supervised machine learning techniques in terms of prediction accuracy and computation load is compared and discussed. The performance evaluation of the developed hybrid modelling approach is shown in Section 6. Lastly, the conclusions are drawn in Section 7.

2. Description of a typical, detached, two-story house in the U.S. and installed AC system

As shown in Fig.1, the reference building being modelled in this research is a typical, single-family, detached house located in Knoxville, Tennessee. It was built in 2013 and is part of a large subdivision of similar homes built around the same time. The 2-story, 223 m^2 (2,400 ft^2) was built to meet the International Energy Conservation Code (IECC) 2006 [60]. The home energy rating score (HERS) value is a metric used to compare the expected energy use of different homes. The building being modelled in this paper achieved a HERS score of 92, which is 8% better than RESNET's reference value of 100 (representing a 2004 IECC built home) [61]. Additional information can be found in [62].

Fig.1. A view of reference house in Tennessee

The house was unoccupied to eliminate the uncertainty and variability introduced with human occupants. However, the sensible and latent heat loads associated with occupancy were simulated using sensible heaters and ultrasonic humidifiers. These were run on a schedule to provide all sensible and latent internal gains (e.g., humans, lighting, cooking, appliance usage) based on the Building America House Simulation Protocol [63].

The overall mean indoor air temperature is defined as the average of measured temperatures from two sensors which are located near the thermostats downstairs and upstairs respectively, as shown in Fig. 2 and Fig. 3.

Fig.2. Thermostat at downstairs

Fig.3. Thermostat at upstairs

The air-conditioning system consists of a variable-speed heat pump, a fan-coil air handler with variable-speed blower, and a zoning system dividing the home into two zones (upstairs and

downstairs). The dampers used to control airflow by the zoning system can modulate their position between the fully-open and fully-closed positions. The position of the damper is reported by the zoning system as an integer between 0 and 15 (0 being fully-closed and 15 being fully-open), and these positions were used to determine the ratio of supply-air delivered to each zone.

3. Natural convection between two floors and developed hybrid modelling approach

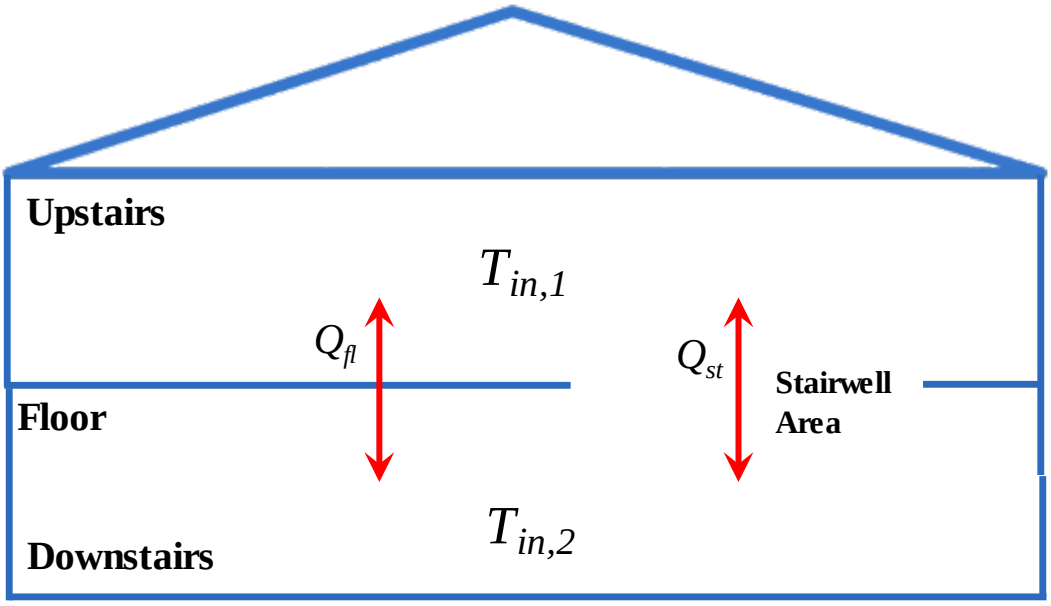

Fig.4 Schematic of heat transfer between upstairs and downstairs in typical detached house in the U.S.

As shown in Fig.4, for prediction of average indoor temperatures in the downstairs ($T_{in.1}$) and upstairs ($T_{in.2}$), both conduction (Q_{fl}), heat transfer through the floor, and convection (Q_{st}) in the open stairwell area need to be calculated. A number of studies have been carried out on the natural convection in a certain space. These studies can be divided into three types of methods:

The first type is the empirical equation for natural convection. The research on the empirical equations of natural convective heat transfer can be divided into large space and limited space. In a large space, the heat transfer can be expressed in the following equations [64, 65]:

$$Gr = \frac{g\alpha\Delta t l^3}{\nu^2} \quad (1)$$

$$Nu = C(GrPr)^n \quad (2)$$

where, *Gr* is Grashof number, g is acceleration of gravity, *Pr* is Prandtl number and *Nu* is Nusselt number (a measure of convection heat transfer at a surface), Δt is the temperature difference to generate natural convection, *C* is coefficient which is relevant to the type of object. *v* is kinematic viscosity, *l* is characteristic length (stair area) α is volume expansion coefficient.

In the limited space, the air flow between two plates was investigated [66, 67]. The empirical equations of natural convective heat transfer were given as follows:

$$\mathrm{Gr} = \frac{g\alpha\Delta t\delta^3}{\nu^2} \tag{3}$$

$$Nu = C(GrPr)^n(H/\delta)^m \tag{4}$$

For a vertical limited space, the equation can be expressed as follows:

$$Nu = 0.197(GrPr)^{\frac{1}{4}}(\frac{H}{\delta})^{-1/9} \tag{5}$$

where, δ is the width of the limited space and *H* is the volume expansion coefficient.

As seen, the natural convection can be calculated through the above empirical equations. However, there are some inadequacies for these equations:

1. They are used for air flow around an object, which is different from the research case in the current study.

2. The solution of convection depends on the temperature difference, which is difficult to achieve in the current study.

The second type is the experimental method. It is realized by carrying out experiments to measure the heat transfer between the downstairs and upstairs. However, it should be noticed that a large number of experiments should be conducted to construct the relationship between the natural convective heat transfer and the influential factors, which is not practical in our research project.

The third type is the computational fluid dynamics (CFD) method. CFD is good at studying indoor air quality, natural ventilation, and stratified ventilation as they are difficult to predict by other methods [68]. Costa investigated the natural convection in the partially divided space, which approximated the second-floor of a house. The air temperature distribution and heat transfer can be calculated [69]. Jin et al [70] adopted a coarse-grid and fast fluid dynamics (FFD) technology to reduce the computing time and get an accurate result. However, the equations used in the CFD method, and even in the simplified FFD method, are very complex. Considerable input data and foreknown knowledge for each single house is needed for this “physical” modelling method, which makes it impractical for application in large-scale building thermal response and power flexibility identification.

In general, obvious natural convection occurs when the indoor air temperature of downstairs is larger than that of upstairs. This dynamic and intermittent heat transfer along with mass transfer is hard to be described by a simplified building thermal model, e.g., RC gray-box model. But

rather, the consideration of this phenomenon would probably lead to increase complexity of the model due to the underlying physics.

Therefore, a black-box modeling method is more practical and promising for our research purpose. Q_{st} is set as a function of the temperature difference between the downstairs and upstairs which is determined by the indoor heat gain difference, as shown in Eq. (6):

$$Q_{st} = F(\Delta T) = F(T_{in,1} - T_{in,2}) = F'\left(\Delta Q_{gain}\right) = F'(\Delta Q_{IHL}, \Delta Q_{AC}, \Delta Q_{solar}, \Delta Q_{wall}) \quad (6)$$

where, ΔQ_{gain} is the heat gain difference between the downstairs and upstairs. ΔQ_{IHL} is the difference of sensible heat gains from indoor heat sources downstairs and upstairs (W). ΔQ_{AC} is the difference of cooling supply downstairs and upstairs (W). ΔQ_{solar} is the difference of solar radiation through windows downstairs and upstairs (W). ΔQ_{wall} is the difference of heat conduction through exterior walls downstairs and upstairs (W).

For the purpose of predicting indoor temperatures, Eq. (6) is further simplified into Eq. (7), in which the temperature difference between the downstairs and upstairs, instead of Q_{st}, is predicted by identifying the relationship between temperature difference and heat gain difference by leveraging supervised machine learning algorithms. This is used in conjunction with a gray-box (i.e., RC model) method that is used to predict the overall mean building indoor temperature (i.e., ($T_{in,1}$+ $T_{in,2}$)/2). Using these two methods together allows for $T_{in,1}$ and $T_{in,2}$ to be determined. So far, a hybrid building thermal modelling approach is formed.

$$\Delta T = T_{in,1} - T_{in,2} = F'\left(\Delta Q_{gain}\right) = F'(\Delta Q_{IHL}, \Delta Q_{AC}, \Delta Q_{solar}, \Delta Q_{wall}) \quad (7)$$

It is worth noting that $T_{in,1}$ and $T_{in,2}$ are the equivalent average temperatures of the downstairs and upstairs respectively, which are measurements from individual sensors (Fig. 2 and 3). It is assumed that the temperature in each floor (zone) is uniform.

In the reference building in this study, for the downstairs and upstairs, the areas of walls and windows in different orientations are almost the same. Therefore, the inputs for machining learning model can be reduced as shown in Eq. (8):

$$\Delta T = T_{in,1} - T_{in,2} = F'\left(\Delta Q_{gain}\right) = F'(\Delta Q_{heater}, \Delta Q_{AC}) \quad (8)$$

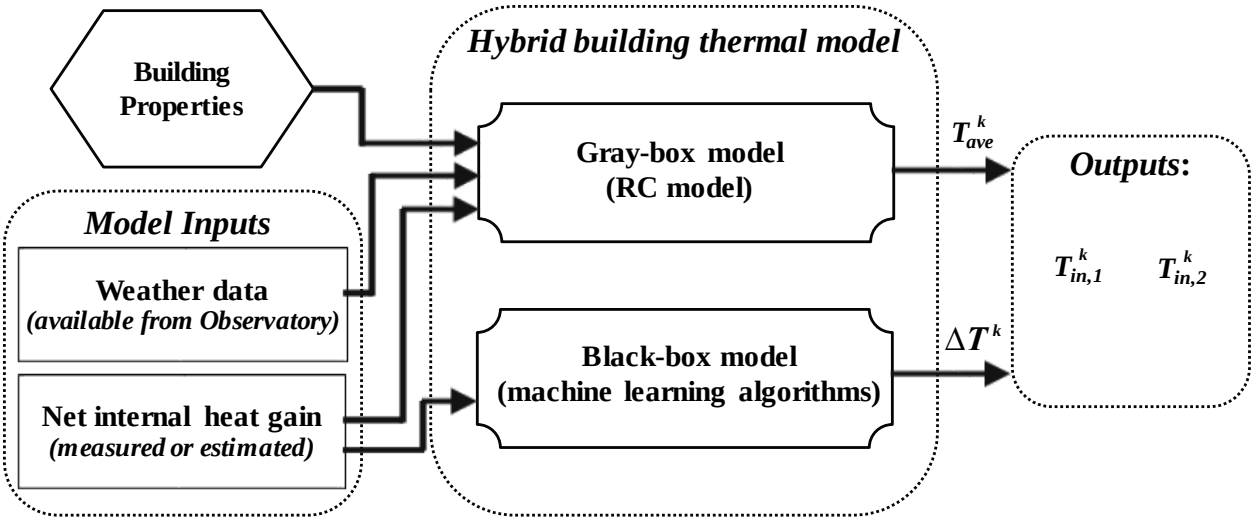

Fig.5 Flowchart of the prediction process of equivalent average temperatures in downstairs and upstairs

As shown in Fig.5, the flowchart exhibits how the indoor temperatures in the downstairs and upstairs are predicted by combining the results from the RC model and the black-box model. For gray-box model, net internal heat gain, weather data, and building properties are used as inputs. The black-box models, in which the supervised machine learning algorithms are leveraged, uses the net internal heat gain (ΔQ_{IHL} and ΔQ_{AC}), as inputs. The superscript k means the kth sampling time.

The details of the RC model and black-box model as well as the corresponding required inputs and building information are elaborated in the following sections.

4. Gray-box model for prediction of average indoor temperature

In this section, the Gray-box model structure is introduced. The training and testing results of the RC model are then shown. Accuracy indices of mean average error (MAE) and root mean squared error (RMSE), are introduced to evaluate the model performance.

4.1 Outline of mathematical model

This gray-box model is used to predict the overall mean building indoor temperature, i.e. ($T_{in,1}$+$T_{in,2}$)/2. which are measured by individual sensors (Fig. 2 and 3). They represent the average temperature in each floor respectively. The schematic of the developed RC building thermal model is shown in Fig. 6.

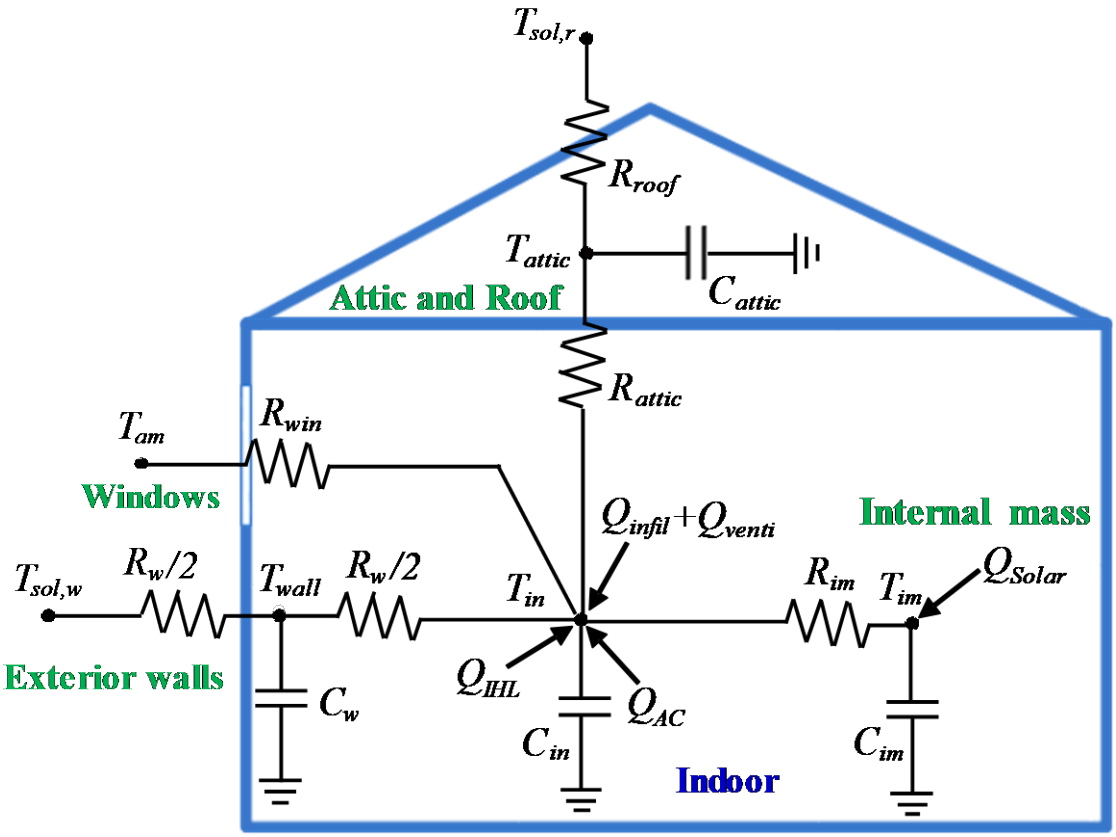

Fig.6. Schematic of the simplified building thermal network model (4R4C)

The heat transfer in the building model is described using the following set of first-order differential equations:

$$C_w \frac{dT_{wall}(t)}{dt} = \frac{T_{sol,w}(t)-T_{wall}(t)}{R_w/2} - \frac{T_{wall}(t)-T_{ave}(t)}{R_w/2} \quad (9)$$

$$C_{in} \frac{dT_{ave}(t)}{dt} = \frac{T_{wall}(t)-T_{ave}(t)}{R_w/2} + \frac{T_{attic}(t)-T_{ave}(t)}{R_{attic}} + \frac{T_{im}(t)-T_{ave}(t)}{R_{im}} + \frac{T_{am}(t)-T_{ave}(t)}{R_{win}} + C_1 Q_{IHL} + C_2 Q_{AC} + Q_{venti} + Q_{infil} \quad (10)$$

$$C_{attic} \frac{dT_{attic}(t)}{dt} = \frac{T_{sol,r}(t)-T_{attic}(t)}{R_{roof}} - \frac{T_{attic}(t)-T_{ave}(t)}{R_{attic}} \quad (11)$$

$$C_{im} \frac{dT_{im}(t)}{dt} = -\frac{T_{im}(t)-T_{ave}(t)}{R_{im}} + C_3 Q_{solar} \quad (12)$$

Where, T_{ave} is the average of $T_{in,1}$ and $T_{in,2}$.

C_w, C_{attic}, C_{im}, and C_{in} are the equivalent overall thermal capacitances of exterior wall, air in attic, internal mass and indoor air respectively. R_w, R_{attic}, R_{roof}, R_{im} and R_{win} are the equivalent overall thermal resistance of exterior walls, attic floor, roof, internal mass and windows respectively. All resistances and capacitances are assumed to be time-invariant.

Building internal mass includes, interior partitions, furniture, etc. It absorbs solar radiation passing through the windows [71]. It is necessary to consider the building internal mass independently since the effect of building internal mass on cooling/heating energy consumption and indoor temperature is significant. The area of internal mass is difficult to calculate and thus assumed to be the floor area although the actual area of internal mass is larger than the floor area.

Q_{IHL} sensible heat gains from indoor heat sources (W), e.g. human, equipment and lighting, which is approximated by adding the sum of circuits, i.e. the total electrical energy use for each electrical circuit in the house, on each level for the separate floor. The energy use of some circuits is not converted to heat inside the home such as the air conditioner, garage outlets, exterior lights, water heater (located in the garage), and the dryer (where most of the heat is vented outside). The power use on these circuits was subtracted from the estimation of the internal heat load. Q_{IHL} is calculated from the following equations:

$$Q_{IHL,1} = W_{house,1} - W_{AC,1} - W_{WH,1} - W_{dryer,1} - W_{garage,1} - W_{ex,l,1} - sum(W_{n,2}) \quad (13)$$

$$Q_{IHL,2} = sum(W_{n,2}) \quad (14)$$

$$Q_{IHL} = Q_{IHL,1} + Q_{IHL,2} \quad (15)$$

where, W is the power consumption (W). The subscripts *house*, *AC*, *WH*, *dryer*, *garage* and *ex,l* mean the whole house, AC, water heater, dryer, garage and exterior lights respectively. The subscripts *1* and *2* mean downstairs and upstairs respectively. The subscript *n* means the total number of circuits. Q_{IHL} is assumed to be known in this study and a model is required to predict it in future studies.

Q_{AC} is the measured total cooling capacity (W) of the AC. For calculation of the separate Q_{AC} for each floor, i.e. $Q_{AC,1}$ and $Q_{AC,2}$, the damper positions, relative duct area, and maximum flow capacity in each duct are considered.

Infiltration and mechanical ventilation loads are calculated based on Eq. (16) and (17). The house being modelled is equipped with a fresh air ventilation system that operates continuously. Therefore, the heat transfer due to ventilation can be calculated based on the mass flow rate of air being exchanged, the specific heat capacity of air, and the temperature difference between the outdoor, fresh air and the indoor air. Estimating the amount of indoor air that is exchanged with the outdoors due to leakage in the building or infiltration is extremely complicated. A simplified approach is used instead, where the infiltration rate is assumed to scale linearly with the wind speed (available from weather data and forecasts). A coefficient, i.e. Coe_w, is used to scale the wind speed to the infiltration rate and is estimated during the training process.

$$Q_{venti} = C_p \cdot V \cdot \rho \cdot (T_{am}(t) - T_{ave}(t)) \quad (16)$$

$$Q_{infil} = C_p \cdot (T_{am}(t) - T_{ave}(t)) \cdot Coe_w \cdot Sp \quad (17)$$

where, V is the mass flow rate of air being exchanged (m^3/s). ρ is the air density (kg/m^3). C_p is the specific heat of air. Sp is the wind speed (m/s).

The solar radiation through window is characterized by Q_{solar} (W) in the following:

$$Q_{solar} = F_{win}(t)I(t)A_{win,tot}SHGC \quad (18)$$

where, I is the direct normal solar irradiance (W/m^2). $A_{win,tot}$ is the total window area (m^2). $SHGC$ is the solar heat gain coefficient of windows. F_{win} is the area-weighted average of view factors of windows with different orientations, which are calculated using a solar calculator spreadsheet developed by NOAA [72]:

$$F_{win}(t) = \frac{\sum_{i=1}^{4} A_{win,i}F_i(t)}{\sum_{i=1}^{4} A_{win,i}} \quad (19)$$

Subscript i indicates the orientations, i.e. east, south, west and north. A_{win} is the general area of window (m^2). F_i is the view factor for windows with different orientations.

One assumption in this research is that the available information is limited in terms of the number of measured points, e.g. only two indoor temperature data measurements available, and general properties of building materials. The effective heating/cooling gain coefficients C_1, C_2 and C_3 are therefore introduced as one main innovation. C_1, C_2 and C_3 are used to adjust Q_{IHL}, Q_{AC} and Q_{solar} for unknown factors. All C_1, C_2 and C_3 are assumed to be unknown and need to be identified by searching algorithm illustrated in Sub-section 4.2.

For Q_{IHL}, the unknown factors include realistic sensible heat ratio (SHR) of internal loads (e.g., cooking often results in electrical energy being converted to both sensible and latent heat) and possible exhausted heat (e.g., range hood operation during cooking can reduce the sensible heat gain). For Q_{AC}, unknown factors include SHR with typical range of 0.6 to 0.8 and the installed inefficiencies such as long refrigerant lines, low indoor airflow, dirty coils, improper refrigerant charge, etc., which are estimated at 10-20%. For Q_{solar}, the unknown factors include window shading such as insect screens, blinds/curtains and external shading.

It is assumed that C_1Q_{IHL} and C_2Q_{AC} are transmitted to indoor air directly by convection and C_3Q_{solar}, is absorbed by internal thermal mass.

The effects of solar radiation on exterior walls and the roof are considered by calculation of $T_{sol,w}$ and $T_{sol,r}$ respectively, which are the sol-air temperatures, shown in the followings:

$$T_{sol,w} = \frac{\alpha_w}{h(t)}F_w(t)I(t) + T_{out} \quad (20)$$

$$T_{sol,r} = \frac{\alpha_r}{h(t)}F_r(t)I(t) + T_{out} \quad (21)$$

where, α is the absorption coefficient of wall and roof. T_{out} is outdoor dry bulb temperature (°C). h is convective heat transfer coefficient of roof and exterior wall surfaces (W/m^2 K), which is calculated by the correlation between h and wind speed developed from ASHRAE Handbook [73]:

$$h = xSp + y \quad (22)$$

where, x and y are the regression coefficients.

F_w and F_r in Eq. (20) and Eq. (21) are the area-weighted averages of view factors of exterior walls and roofs with different orientations.

where, A_{wall} is the general area of each wall (m^2). F_i is the view factor for walls with different orientations. Subscript j indicates the orientations, i.e. north and south, of roof.

$$F_w(t) = \frac{\sum_{i=1}^{4} A_{wall,i} F_i(t)}{\sum_{i=1}^{4} A_{wall,i}} \tag{23}$$

$$F_r(t) = \frac{\sum_{j=1}^{2} F_j(t)}{2} \tag{24}$$

A comparison among F_r, F_w and F_{win} in Sep 1, 2012 is shown in Fig.7 for illustration purpose.

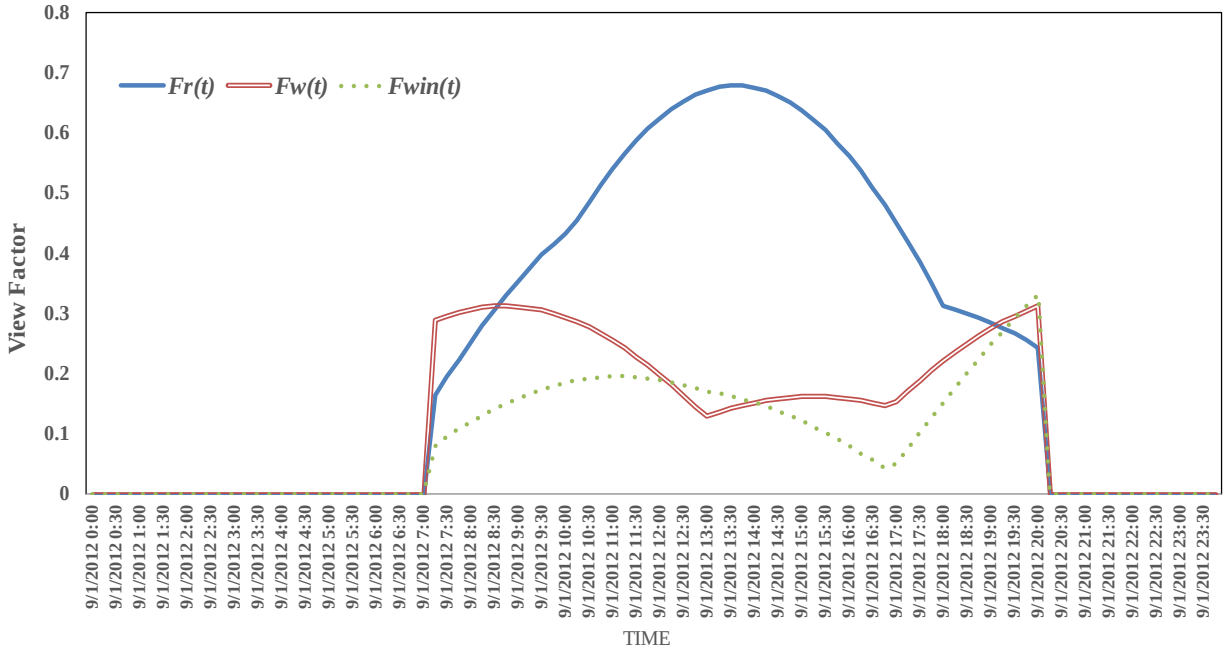

Fig.7 Comparison of view factors in a typical day in September

In summary, three main data sources are required for RC model development. The weather data are assumed to be available from a weather station directly. The building properties are available from building plans or general estimations for calculation of solar radiation and sol-air temperature. The measured or estimated data include the whole house power, HVAC power, water heater power, dryer power and power use in unconditional spaces.

4.2. Parameters identifications

The searching process for optimal values of the undetermined parameters in this model is a nonlinear optimization process. Given a set of parameters, the gray-box model can predict the indoor air temperature (T_{in}) profile. An objective function is used to evaluate the fitness between

the predicted results and measured data collected from the reference building during the training period. The objective function *J* of such optimization is to minimize the integrated RMSE, as defined in Eq. (25):

$$J(R_{w,}R_{roof,}R_{attic,}R_{im,}R_{win,}C_{w,}C_{im,}C_{attic,}C_{in,}C_1, C_2, C_3, R_{win}, Coe_w) = \sqrt{\frac{\sum_{k=1}^{N}(T_{ave,act}-T_{ave,simu})^2}{N-1}} \quad (25)$$

where, $T_{ave,act}$ is the measured overall mean building indoor dry bulb temperature. $T_{ave,simu}$ is the result from the model. The parameters are identified by particle swarm optimization (PSO) method. PSO is a computational method that optimizes a problem by iteratively trying to improve a candidate solution with regards to a given measure of quality. It solves a problem by having a population of candidate solutions and moving these particles around in the search-space according to simple mathematical formulae over the particle's position and velocity. The searching ranges for C_w, C_{attic} and all the R are referred to the recommended parameters values of buildings in zone 4 from IECC [60]. C_{im} can be assumed to a be in the range of 100 to 450 (KJ/K m^2) [74]. The lower and upper limits for each R and C are 1/3 and 3 times of the estimated values respectively. The ranges of C_1, C_2 and C_3 are estimated based on experiences.

The model development is written and the PSO package is imported in RStudio [75], which is a free and open-source integrated development environment (IDE) for R.

4.3. Training and testing results of the RC model

In this sub-section, the data collected from the reference building in different consecutive time periods with various operation conditions, e.g. different schedules of AC indoor temperature set-points, as well as different outdoor weather conditions are used for training and validating the RC model. The data collected from 4/21/2017 to 5/16/2017 (25 consecutive days) are used for training section and the data collected from 5/29/2017 to 6/14/2017 (17 consecutive days) are used for testing.

The scheduled AC and heater set-points are listed in Table 1.

Table 1 The scheduled AC and heater set-points for RC model development

Day of the week	Time	heater/AC Set-points
Monday	8:00 am to 4:00 pm	18.3˚C/26.7˚C
	4:00 pm to 8:00 am	21.7˚C/24.4˚C
Tuesday	8:00 am to 5:00 pm	18.3˚C/26.7˚C
	5:00 pm to 8:00 am	20.6˚C/23.3˚C
Wednesday	8:00 am to 5:00 pm	18.3˚C/26.7˚C
	5:00 pm to 8:00 am	19.4˚C/25.6˚C
Thursday	8:00 am to 3:00 pm	19.4˚C/25.6˚C
	3:00 pm to 8:00 am	21.7˚C/24.4˚C
Friday	8:00 am to 6:00 pm	19.4˚C/25.6˚C
	6:00 pm to 8:00 am	20.6˚C/23.3˚C
Saturday	Whole day	20.6˚C/23.3˚C

Sunday	8:00 am to 10:00 pm	21.7˚C/24.4˚C
	10:00 pm to 8:00 am	20.6˚C/23.3˚C

T_{act} is the measured overall mean indoor air temperature and T_{RC} is the indoor temperature from the model for a 24-hour prediction horizon. The resulting parameters identified by PSO are: R_w = 0.0134 K/W, R_{attic} = 0.0235 K/W, R_{roof} = 0.00156 K/W, R_{im} = 0.00171 K/W, C_w = 10,383,364 J/K, R_{roof} = 0.00156 K/W, C_{attic} = 704,168 J/K, C_{im} = 23,396,403 J/K, C_{in} = 8,665,588 J/K, C_1 = 0.691, C_2 = 0.784, C_3 = 0.1, R_{win}= 0.021 K/W and $WinS$ = 0.01.

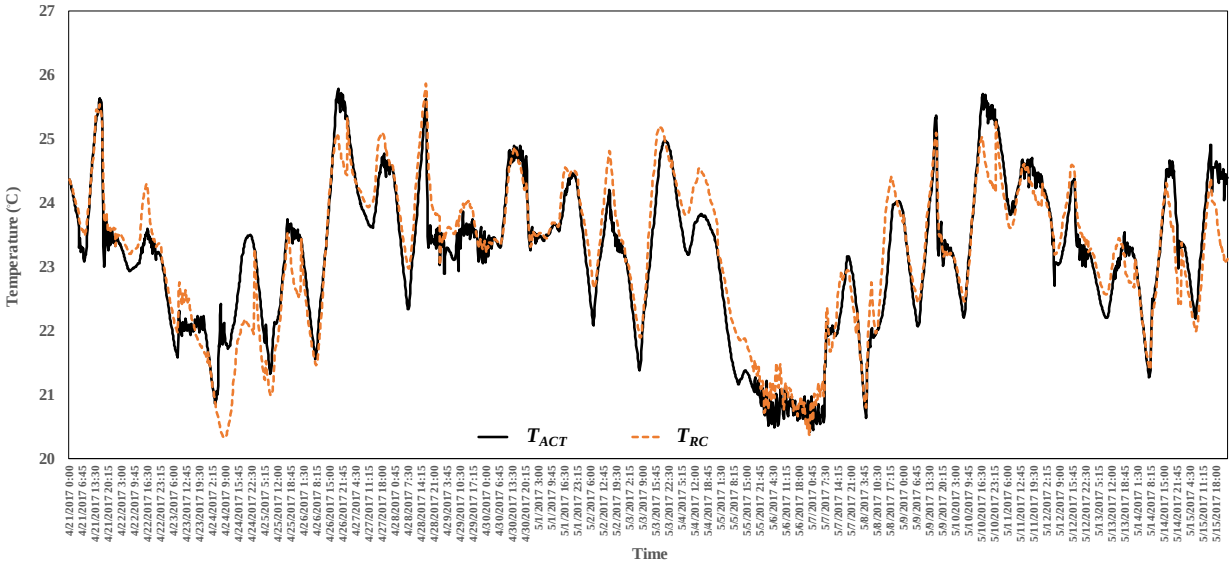

Fig.7 RC model training results from April 21 to May 16, 2017

The data collected from 5/29/2017 12 a.m. to 6/14/2017 12 p.m. are used for testing. The results are shown in Fig. 8.

To quantify the deviations of the predicted data from the measured data in both training session and testing sessions, two indices are used to evaluate the deviations: MAE and RMSE:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|T_{ACT}(t) - T_{RC}(t)| \tag{26}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(T_{ACT}(t) - T_{RC}(t))^2} \tag{27}$$

Table 2 Accuracy indices of the developed RC model

Time	Training/Testing	MAE	RMSE
Apr 21 to May 16, 2017	Training	0.343˚C	0.456˚C
May 29 to Jun 14, 2017	Testing	0.499˚C	0.619˚C

Table 2 lists the two accuracy indices of the developed model in training and testing sessions. It can be found that the developed RC model has satisfactory performance in prediction of the overall mean building indoor temperature under different scenarios.

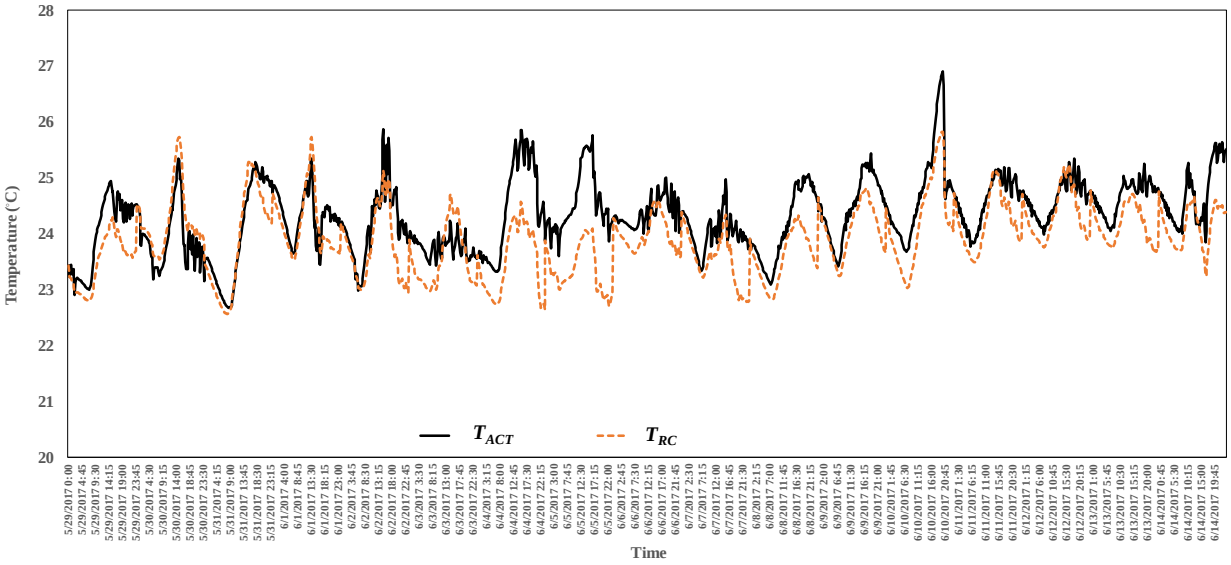

Fig.8 RC model testing results from May 29 to Jun 14, 2017

5. Black-box models for predicting temperature differences between downstairs and upstairs.

This section first introduces seven supervised machine learning algorithms. The black-box models are then built based on these algorithms to predict the temperature difference between the downstairs and upstairs. The prediction performance of these algorithms is shown and compared. Both MAE and RMSE are introduced to evaluate the prediction accuracy.

5.1 Description of seven supervised machine learning algorithms

Seven predictive methods are used for model development, i.e., Generalized Linear Models (GLM), support vector regression (SVR), artificial neural network (ANN), random forests (RF), gradient boosting trees (GBM), extreme gradient boosting (XGB) and light gradient boosting (LGB). These methods are selected based on their performance in previous studies and potentials in describing complex relationships. The first method, i.e., GLM, is used as the performance benchmark, as it is only capable of capturing linear relationships. The other six methods are suitable for discovering the underlying nonlinear relationships between model inputs and outputs. Four of them, i.e., SVR, ANN, RF and GBM, are relatively mature solutions and their performance has been validated in previous studies [43]. The other two methods, i.e., XGB and LGB, are relatively new techniques recently developed in the machine learning community [76, 63]. XGB and LGB are improved versions of gradient boosting trees. They can better tackle the overfitting problem with higher computation efficiency [77].

5.2 Training and testing results of the black box models

In this sub-section, the data collected from the reference building in different three consecutive time periods with various operation conditions, e.g. different schedules of AC indoor temperature set-points, are used to train and validate the black-box model in which seven supervised machine learning algorithms are leveraged.

The data collected from 7/15/2017 00:00 to 7/20/2017 23:45 and 6/30/2017 00:00 to 7/7/2017 23:45 are used for model training. The data collected from 7/25/2017 15:05 to 7/30/2017 23:05 are used for model testing. The entire dataset is accordingly divided into training and testing data with proportions of around 70% and 30% respectively.

In the Fig 9 and 10, the curves marked as "ΔT_{act}" are the measured temperature differences between the downstairs and upstairs. ΔT_{GLM}, ΔT_{ANN}, ΔT_{SVR}, ΔT_{RF}, ΔT_{XGB}, ΔT_{LGB} and ΔT_{GBM} are the training results from the methods of GLM, ANN, SVR, RF, XGB, LGB and GBM respectively.

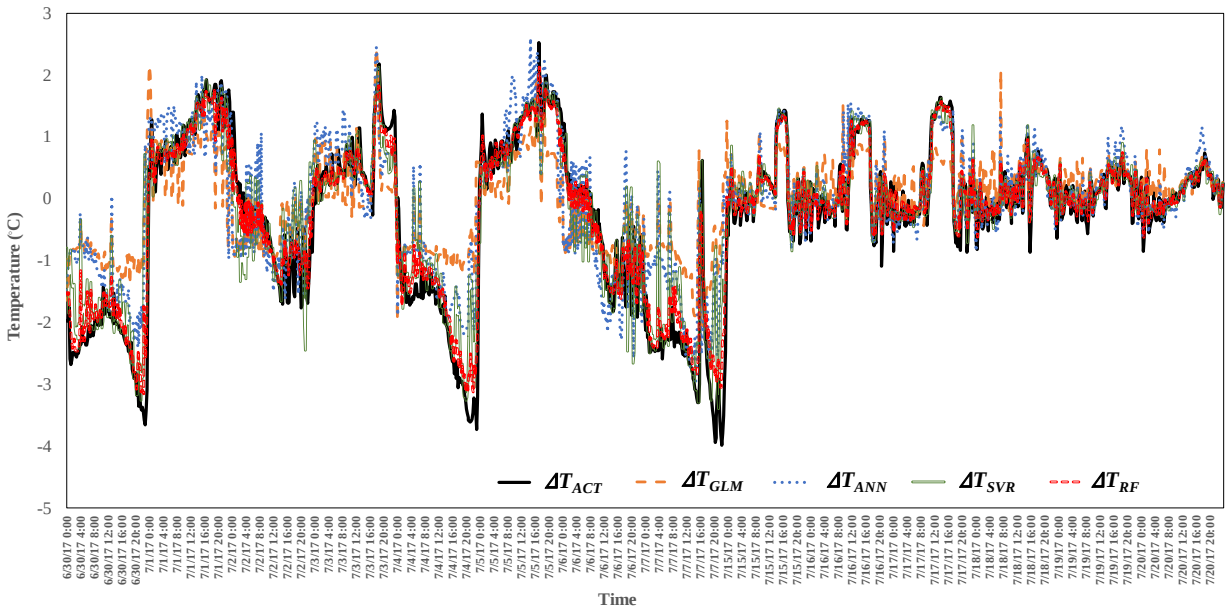

Fig. 9. Training results from black-box models (1 of 2)

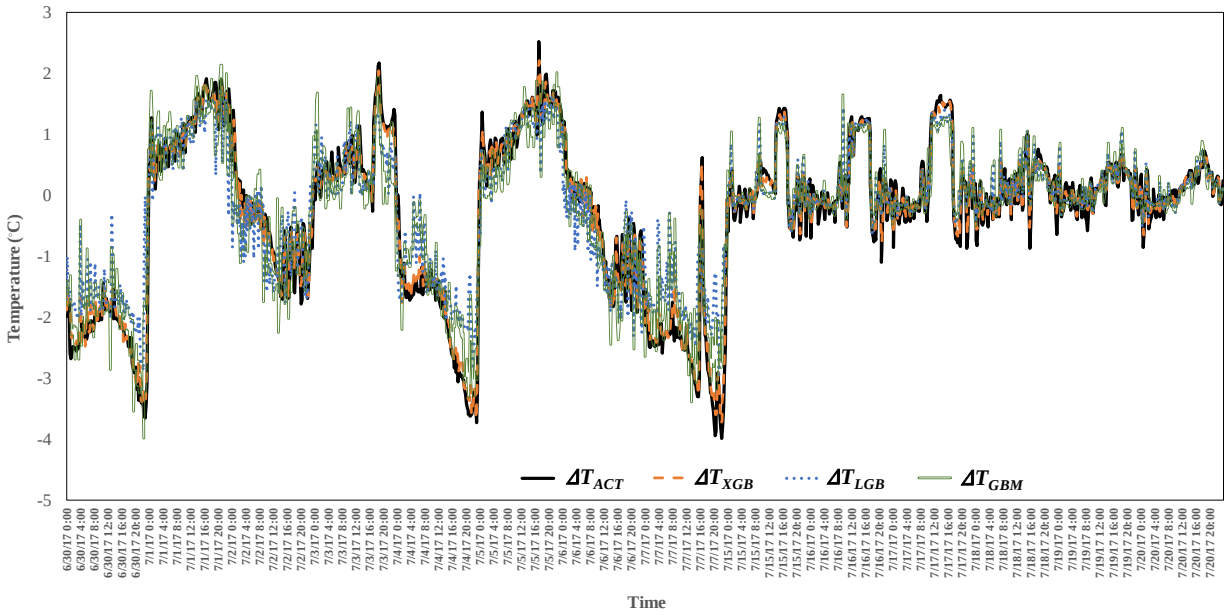

Fig. 10. Training results from black-box model (2 of 2)

The corresponding respective AC indoor temperature set-points for downstairs and upstairs are also shown in Fig 11 and Fig 12. The curves marked as "$T_{set,1}$" and "$T_{set,2}$" are AC set-points of downstairs and upstairs respectively. The set-points of heaters are constants as 18.3˚C.

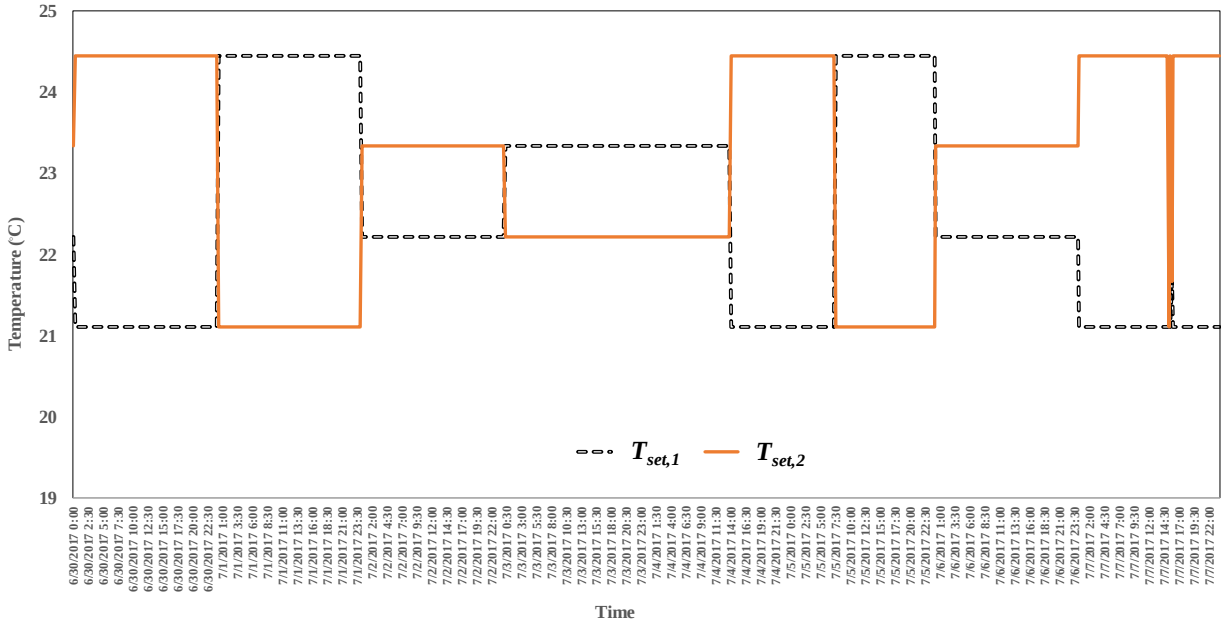

Fig. 11. The indoor temperature setpoints in the training period (6/30/2017 to 7/7/2017)

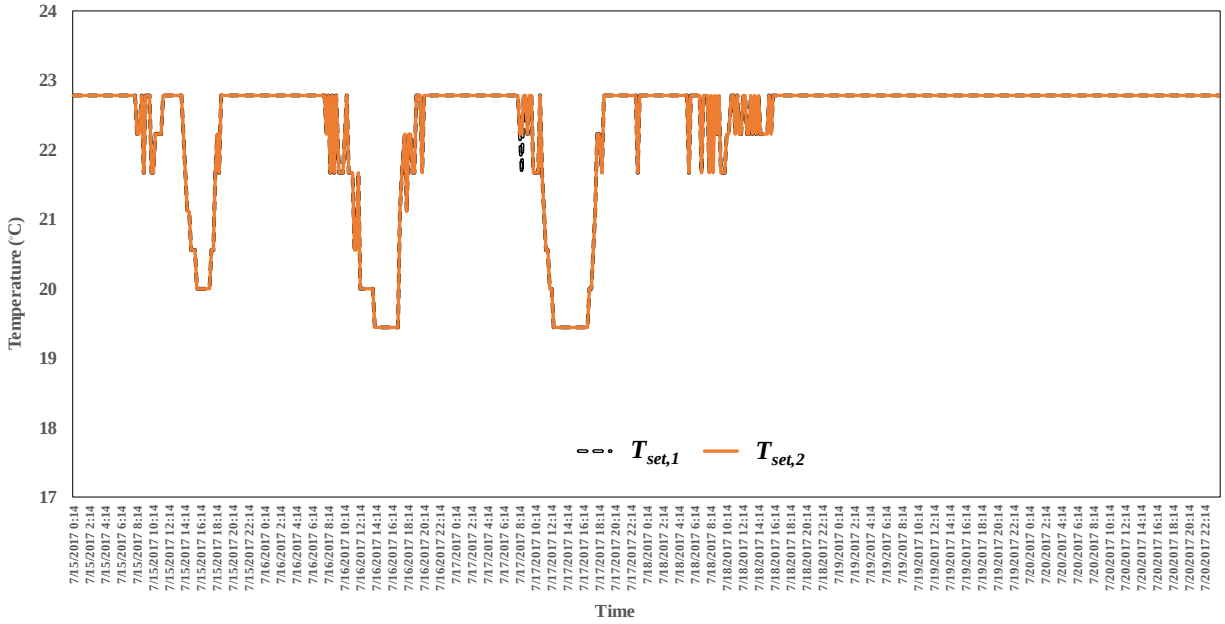

Fig. 12. The indoor temperature setpoints in the training period (7/15/2017 to 7/20/2017)

To ensure the model performance, 3-fold cross-validation is used for parameter optimization. Two parameters are optimized for developing SVR models with a radial basis function kernel, i.e., the complexity parameter *C* and the smoothing parameters *sigma*. The higher the *C* value, the more complex the SVR model. A large *C* may lead to more accurate training performance. However, it may also cause the over-fitting problem. The *sigma* controls the shape of the decision boundary, i.e., the larger the *sigma*, the more flexible and smooth the decision boundary. A three-layer architecture (i.e., one input layer, one hidden layer and one output layer) is used to develop artificial neural network models. The number of neurons at the hidden layer is optimized through cross-validation. In terms of the other four tree-based methods, the number of trees is defined as 100 for comparison purposes. The number of input variables considered for each tree splitting is the main optimization target for RF models. Two parameters, i.e., the interaction level and shrinkage, are optimized for GBM models. The maximum tree depth and the learning rate are optimized for XGB and LGB models. The optimization result is summarized in Table 3.

In the Fig 13 and 14, the testing results from the seven different methods are shown. The corresponding respective AC indoor temperature set-points for downstairs and upstairs are also shown in Fig 15. The set-points of heaters are constants as 18.3˚C.

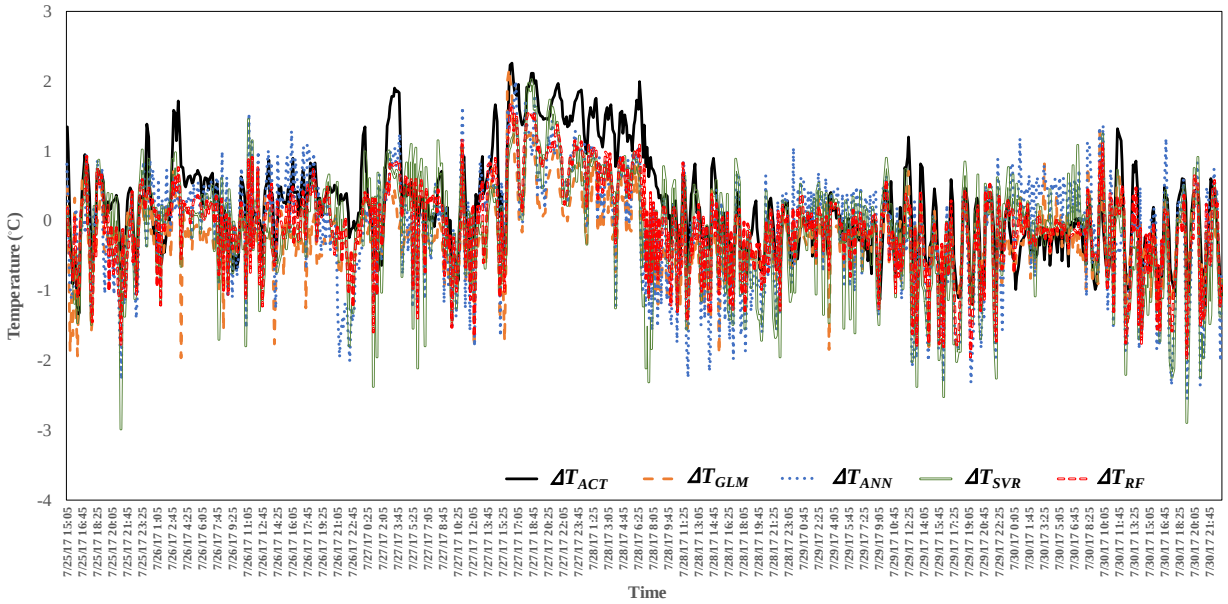

Fig. 13. Testing results from black-box models (1 of 2)

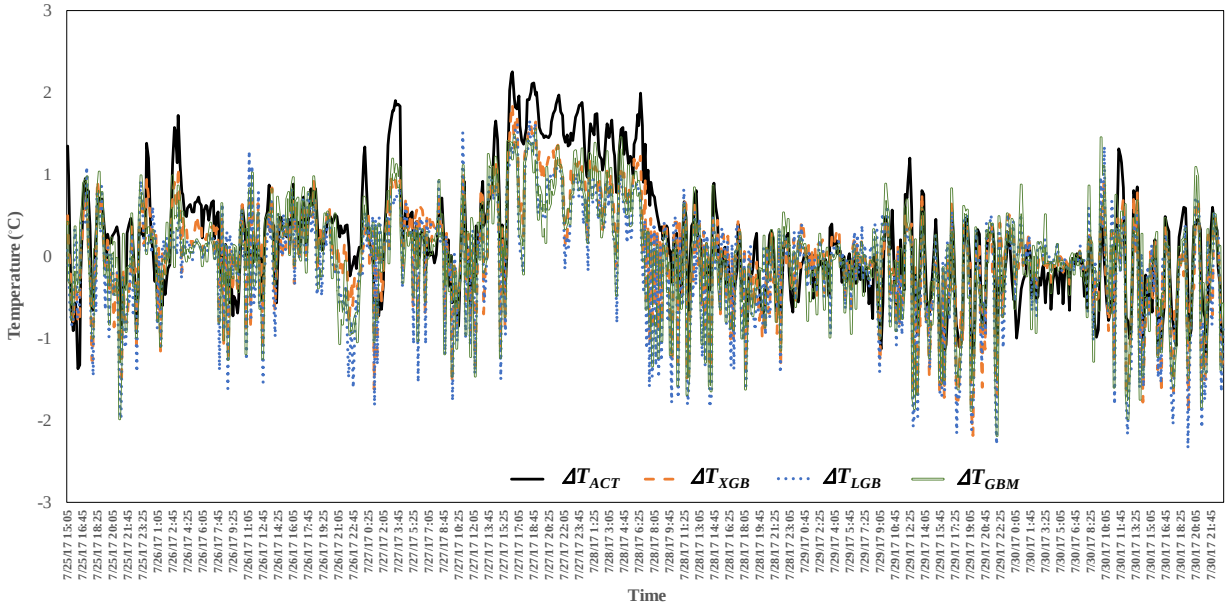

Fig. 14. Testing results from black-box models (2 of 2)

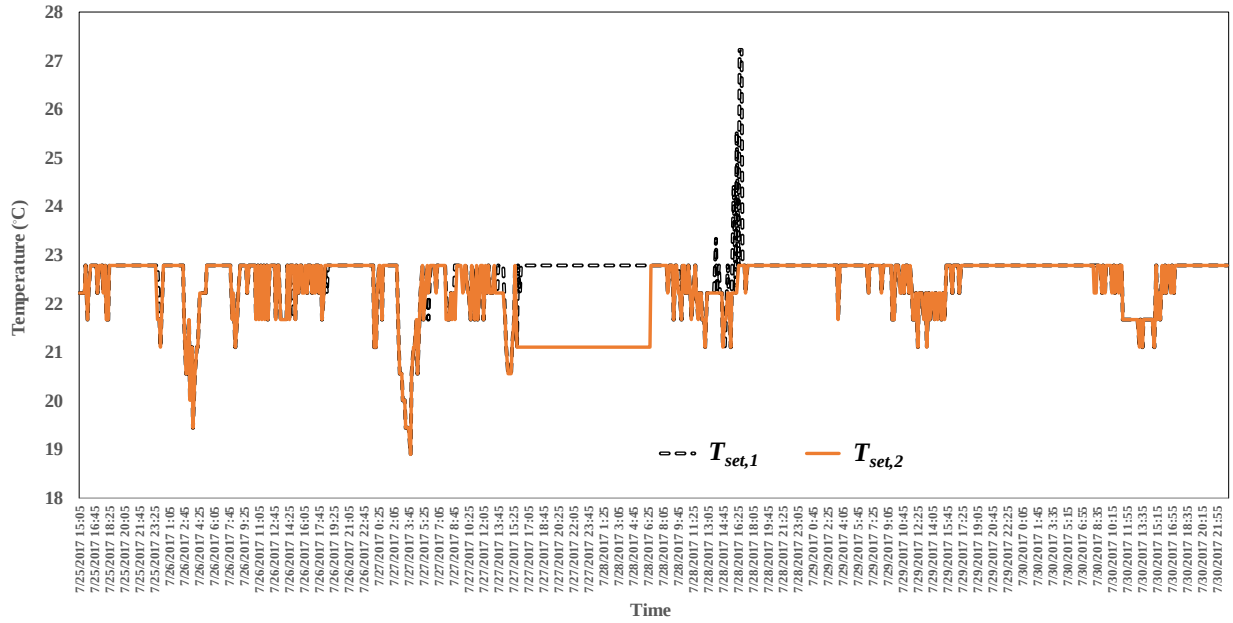

Fig. 15. The indoor temperature setpoints in the training period (7/25/2017 to 7/30/2017)

Table 3 Parameter optimization results

Methods	Parameters	Values
ANN	No. of hidden neurons	20.00
SVR	C	10.00
	Sigma	0.23
RF	No. of candidate splitting variables	2.00
GBM	Interaction	3.00
	Shrinkage	0.10
XGB	Learning rate	0.10
	Max tree depth	6.00
LGB	Learning rate	0.10
	Max tree depth	7.00

The prediction performance of each method is evaluated based on their prediction accuracy and the computation load. Table-4 summarizes the predication accuracy on the testing data and the computation time for model development. The prediction accuracy is also evaluated using RMSE and MAE.

Table 4 lists the accuracy matrices of different methods. In terms of RMSE, The ANN and GBM have relatively poor performance when compared with other nonlinear methods. In terms of the computation time, GLM and ANN result in the shortest and longest computation time respectively. This is in accordance with domain expertise, as the computation time is

proportional to the complexity of model architectures and the number of model parameters. Two recently developed tree-based methods, i.e., LGB and XGB, have the smallest computation time among nonlinear methods. Both methods adopt a parallel way to tree development. In addition, LGB uses histogram-based algorithm to handle numeric variables and therefore, the resulting training speed is faster.

Table 4 Prediction performance of seven black-box models on testing data sets

Methods	MAE	RMSE	Computation time (s)
GLM	0.547°C	0.693°C	0.003
ANN	0.625°C	0.768°C	6.994
SVR	0.541°C	0.701°C	0.260
RF	0.438°C	0.552°C	0.261
GBM	0.448°C	0.575°C	0.338
XGB	0.392°C	0.499°C	0.207
LGB	0.433°C	0.551°C	0.198

6. Performance evaluation of the developed hybrid modeling approach

This section shows the overall prediction results of the temperatures in respective floors are shown. The performance of the developed hybrid modelling approach is then evaluated.

The RC model is used to predict the overall mean temperature in the black-box model testing period (7/25/2017 15:05 to 7/30/2017 23:05). The results are shown in Fig. 16. The MAE and RMSW are 0.435°C and 0.56°C respectively.

The final results, i.e. the predicted average temperatures in respective floors, produced by RC model plus different black-box modeling methods are then calculated. For the sake of clarity, only the results from RC+ XGB are shown in the Fig. 17. $T_{1,act}$ and $T_{2,act}$ are the measured downstairs and upstairs temperatures respectively. $T_{1,XGB}$ and $T_{2,XGB}$ are the predicted downstairs and upstairs temperature respectively. It is obvious that the predicted temperatures are quite closed to the real ones.

The detailed accuracy indices of the developed hybrid modeling approach are listed in Table 5. The results from RC+XGB have the best accuracy. The average MAE and RMSE of RC+XGB are 0.510 and 0.641 respectively.

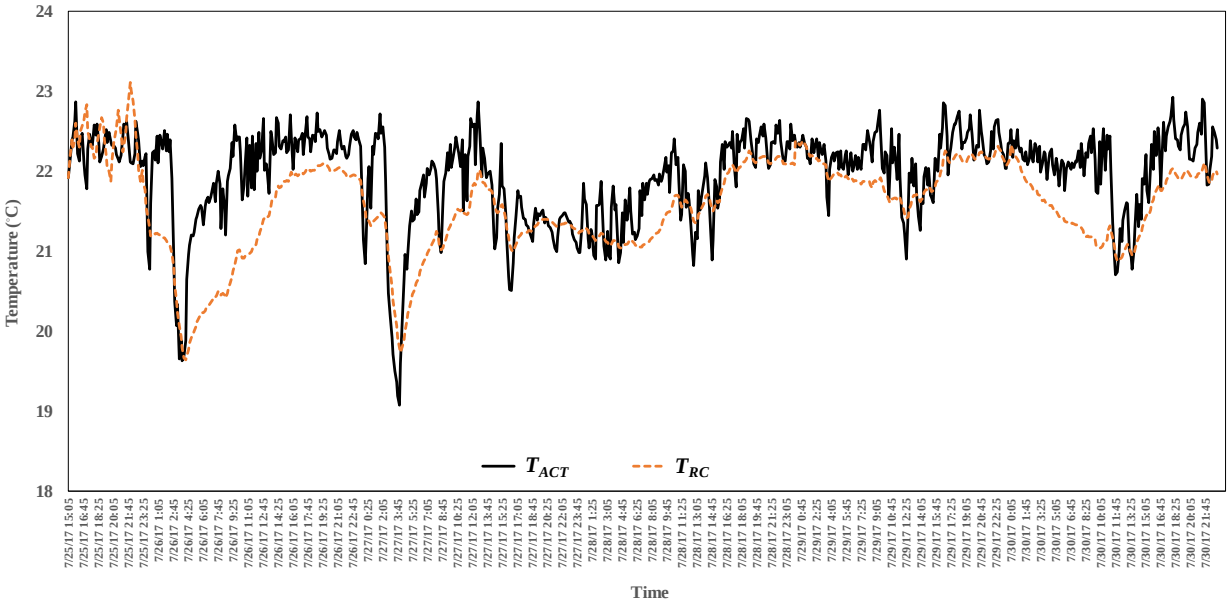

Fig. 16. The predicted overall mean indoor temperature by RC model (7/25/2017 to 7/30/2017)

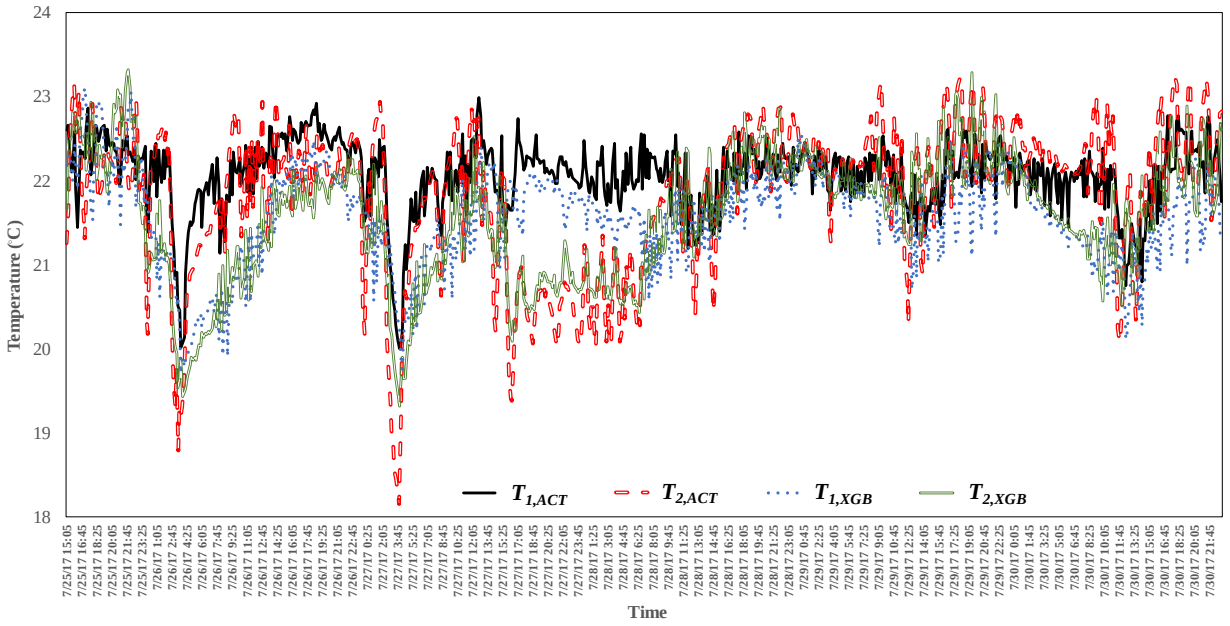

Fig. 18. The predicted indoor air temperatures in respective floors by XGB+RC

Table 5 Overall prediction performance of developed hybrid modeling approach

Gray-box method	**Black-box methods**	Predicted average temperature in respective floors	**MAE**	**RMSE**
RC	GLM	T_1	0.629	0.756
		T_2	0.475	0.605
	ANN	T_1	0.584	0.724

	T_2	0.550	0.684
SVR	T_1	0.592	0.724
	T_2	0.507	0.647
RF	T_1	0.553	0.676
	T_2	0.491	0.626
GBM	T_1	0.537	0.698
	T_2	0.509	0.642
XGB	T_1	0.518	0.643
	T_2	0.502	0.638
LGB	T_1	0.563	0.698
	T_2	0.523	0.660

7. Conclusion

In typical existing detached two-story houses in the U.S., zoned AC systems (either through the use of dampers or two independent ACs) are used to achieve indoor thermal control in both the downstairs and upstairs. The ability to predict the respective indoor air temperatures of both the upstairs and downstairs is required to realize practical and efficient AC control. The traditional gray-box modeling method may not be able to predict the respective temperatures directly, since it is difficult to specify a comprehensive gray-box model to take into account the air mass movement and the corresponding heat transfer in the stair area due to natural convection.

This inspires us to develop a creative and efficient hybrid modeling approach to realize accurate prediction of the indoor air temperatures of both the upstairs and downstairs. In this hybrid method, the developed RC model is used to predict the overall mean indoor air temperature and the data-driven model is used to predict the temperature difference between the downstairs and upstairs. The data from an in-situ typical two-story house are used to validate the proposed approach.

Based on the testing results, the following conclusions can be drawn:

1. Two accuracy indices showed that RC model has satisfactory performance, i.e. 0.499°C of MAE and 0.619°C of RMSE, in prediction of the overall mean building indoor temperature under different scenarios for a 24-hour horizon by using around 3 weeks' data for training section.
2. In terms of the testing results, the adopted seven black-box methods/supervised machine learning techniques have different prediction performance. The ANN and GBM have relatively poor performance when compared with other nonlinear methods. In terms of the computation time, ANN result in the longest computation time. Two recently developed tree-based methods, i.e., XGB and RF, have better accuracies than other methods. For instance, the MAE and RMSE of XGB are 0.399°C and 0.499°C respectively.

3. The developed hybrid approach has excellent performance in prediction of average temperatures in downstairs and upstairs. For instance, the average MAE and RMSE of RC+XGB are 0.510 and 0.641 respectively.

In general, this research provides a practicable, easy-to-use and effective solution for prediction and control of the interactive indoor air temperatures in downstairs and upstairs of two-story houses with open stairwell area. The resulted models are simpler, thus computationally tractable for calculation of the optimal control, e.g. MPC, for large system. It is worth noticing that some features in the developed approach can be promoted and directly applied in modelling of other houses in U.S.: 1. The first-order RC model structure which considers the non-homogeneous heat transfer features of exterior wall, roof and internal mass separately. 2. the identification method of the view factors in RC model. 3. The available data of AC cooling supply and sensible heat outputs from human and/or heaters in downstairs and upstairs.

The research is part of a project named “Virtual battery-based characterization and control of flexible building”. The overall goal of the project is to optimally control building loads to behave as virtual storage to provide grid services, integrate more renewable generation and improve building operational efficiency while meeting the needs of building occupants. This work facilitates that aim by developing robust and accurate building thermal modeling method.

Acknowledgments

The research presented in this paper was funded by the United States Department of Energy, Energy Efficiency and Renewable Energy Office, Building Technology Office under the grid modernization program.

References

[1] DOE. Buildings energy data book tech rep. U.S. Department of Energy; 2011.

[2] Cole, W.J., Powell, K.M., Hale, E.T., Edgar, T.F. Reduced-order residential home modeling for model predictive control. Energy and Buildings 2014; 74: 69-77.

[3] Wattles P. ERCOT demand response overview & status report. InProc. AMIT-DSWG Workshop AMI's Next Front. Demand Response 2011 Aug 30 (pp. 1-25).

[4] Motegi N et al. Introduction to commercial building control strategies and techniques for demand response. Berkeley: Lawrence Berkeley National Laboratory; 2005. LBNL-59975.

[5] Cui, B., Gao, D.C., Xiao, F. and Wang, S., 2017. Model-based optimal design of active cool thermal energy storage for maximal life-cycle cost saving from demand management in commercial buildings. Applied Energy, 201:382-396.

[6] Sun, Y., Wang, S., Xiao, F. and Gao, D., 2013. Peak load shifting control using different cold thermal energy storage facilities in commercial buildings: a review. Energy conversion and management, 71, pp.101-114.

[7] Taylor, J.A. and Mathieu, J.L., 2014. Index policies for demand response. IEEE Transactions on Power Systems, 29(3), pp.1287-1295.

[8] Luthander, R., Widén, J., Nilsson, D. and Palm, J., 2015. Photovoltaic self-consumption in buildings: A review. *Applied Energy*, *142*, pp.80-94.

[9] Cao, S. and Sirén, K., 2014. Impact of simulation time-resolution on the matching of PV production and household electric demand. *Applied Energy*, *128*, pp.192-208.

[10] Dong, J., Djouadi, S.M., Kuruganti, T. and Olama, M.M., 2017, August. Augmented optimal control for buildings under high penetration of solar photovoltaic generation. In *Control Technology and Applications (CCTA), 2017 IEEE Conference on* (pp. 2158-2163). IEEE

[11] Sharma, I., Dong, J., Malikopoulos, A.A., Street, M., Ostrowski, J., Kuruganti, T. and Jackson, R., 2016. A modeling framework for optimal energy management of a residential building. Energy and Buildings, 130, pp.55-63.

[12] Alimohammadisagvand, B., Jokisalo, J., Kilpeläinen, S., Ali, M. and Sirén, K., 2016. Cost-optimal thermal energy storage system for a residential building with heat pump heating and demand response control. *Applied Energy*, *174*, pp.275-287.

[13] Patteeuw, D., Reynders, G., Bruninx, K., Protopapadaki, C., Delarue, E., D'haeseleer, W., Saelens, D. and Helsen, L., 2015. CO2-abatement cost of residential heat pumps with active demand response: demand-and supply-side effects. *Applied Energy*, *156*, pp.490-501.

[14] Álvarez, S., Cabeza, L.F., Ruiz-Pardo, A., Castell, A. and Tenorio, J.A., 2013. Building integration of PCM for natural cooling of buildings. *Applied energy*, *109*, pp.514-522.

[15] Fiorentini, M., Cooper, P. and Ma, Z., 2015. Development and optimization of an innovative HVAC system with integrated PVT and PCM thermal storage for a net-zero energy retrofitted house. *Energy and Buildings*, *94*, pp.21-32.

[16] Pavlak, G.S., Henze, G.P. and Cushing, V.J., 2015. Evaluating synergistic effect of optimally controlling commercial building thermal mass portfolios. *Energy*, *84*, pp.161-176.

[17] Wang, S. and Xu, X., 2006. Parameter estimation of internal thermal mass of building dynamic models using genetic algorithm. *Energy conversion and management*, *47*(13), pp.1927-1941.

[18] Zhou, Q., Wang, S., Xu, X. and Xiao, F., 2008. A grey-box model of next-day building thermal load prediction for energy-efficient control. *International Journal of Energy Research*, *32*(15), pp.1418-1431.

[19] Yin, R., Xu, P., Piette, M.A. and Kiliccote, S., 2010. Study on Auto-DR and pre-cooling of commercial buildings with thermal mass in California. *Energy and Buildings*, *42*(7), pp.967-975.

[20] Lee, K.H. and Braun, J.E., 2008. Model-based demand-limiting control of building thermal mass. *Building and Environment*, *43*(10), pp.1633-1646.

[21] Braun, J.E., Montgomery, K.W. and Chaturvedi, N., 2001. Evaluating the performance of building thermal mass control strategies. *HVAC&R Research*, *7*(4), pp.403-428.

[22] Hurtado, L.A., Rhodes, J.D., Nguyen, P.H., Kamphuis, I.G. and Webber, M.E., 2017. Quantifying demand flexibility based on structural thermal storage and comfort management of non-residential buildings: A comparison between hot and cold climate zones. Applied Energy, 195, pp.1047-1054.

[23] Xue, X., Wang, S., Yan, C. and Cui, B., 2015. A fast chiller power demand response control strategy for buildings connected to smart grid. *Applied Energy*, *137*, pp.77-87.

[24] American Society of Heating, Refrigerating and Air Conditioning Engineers, Inc., ASHRAE Handbook-Fundamentals, D&R International, Ltd, 2013.

[25] Coakley, D., Raftery, P. and Keane, M., 2014. A review of methods to match building energy simulation models to measured data. *Renewable and sustainable energy reviews*, *37*, pp.123-141.

[26] Braun, J.E. and Chaturvedi, N., 2002. An inverse gray-box model for transient building load prediction. *HVAC&R Research*, *8*(1), pp.73-99.

[27] Dong, B., Li, Z., Rahman, S.M. and Vega, R., 2016. A hybrid model approach for forecasting future residential electricity consumption. Energy and Buildings, 117, pp.341-351.

[28] Bacher, P. and Madsen, H. Identifying suitable models for the heat dynamics of buildings. Energy and Buildings 2011; 43(7): 1511-1522.

[29] Kim, D., Cai, J., Ariyur, K.B. and Braun, J.E., 2016. System identification for building thermal systems under the presence of unmeasured disturbances in closed loop operation: Lumped disturbance modeling approach. Building and Environment, 107, pp.169-180.

[30] F Belić, Ž Hocenski, D Slišković. Thermal modeling of buildings with RC method and parameters estimation. 2016 International Conference on Smart Systems and Technologies (SST); 12-14 Oct 2016, page:19-25.

[31] C Lombard, EH Mathews. Efficient, steady state solution of a time variable RC network, for building thermal analysis. Building and Environment 1992; 27(3):279-287.

[32] Wang, S. and Xu, X., 2006. Simplified building model for transient thermal performance estimation using GA-based parameter identification. *International Journal of Thermal Sciences*, *45*(4), pp.419-432.

[33] Fiorentini, M., Wall, J., Ma, Z., Braslavsky, J.H. and Cooper, P., 2017. Hybrid model predictive control of a residential HVAC system with on-site thermal energy generation and storage. Applied Energy, 187, pp.465-479.

[35] Goyal, S. and Barooah, P., 2012. A method for model-reduction of non-linear thermal dynamics of multi-zone buildings. *Energy and Buildings*, *47*, pp.332-340.

[36] Henze, G.P., Pavlak, G.S., Florita, A.R., Dodier, R.H. and Hirsch, A.I., 2015. An energy signal tool for decision support in building energy systems. Applied Energy, 138, pp.51-70.

[39] Zhu, Q., Xu, X., Gao, J. and Xiao, F., 2015. A semi-dynamic model of active pipe-embedded building envelope for thermal performance evaluation. International Journal of Thermal Sciences, 88: 170-179.

[40] Zhu, N., Hu, P. and Xu, L., 2013. A simplified dynamic model of double layers shape-stabilized phase change materials wallboards. Energy and Buildings, 67, pp.508-516.

[41] Cui, B., Wang, S., Yan, C. and Xue, X., 2015. Evaluation of a fast power demand response strategy using active and passive building cold storages for smart grid applications. Energy Conversion and Management, 102, pp.227-238.

[42] Cui, B., Gao, D.C., Wang, S. and Xue, X., 2015. Effectiveness and life-cycle cost-benefit analysis of active cold storages for building demand management for smart grid applications. Applied energy, 147:523-535.

[43] Zhao, H.X., Magoules, F., 2012. A review on the prediction of building energy consumption. Renewable and Sustainable Energy Reviews, 16, pp. 3586-3592.

[44] Wei, Y.X., Zhang, X.X., Shi, Y., Xia, L., Pan, S., Wu, J.S., Han, M.J., Zhao, X.Y., 2018. A review of data-driven approaches for prediction and classification of building energy consumption. Renewable and Sustainable Energy Reviews, 82, pp. 1027-1047.

[45] Amasyali, K., El-Gohary, N.M., 2018. A review of data-driven building energy consumption prediction studies. Renewable and Sustainable Energy Reviews, 81, pp. 1192-1205.

[46] Ghiaus, C., 2006. Experimental estimation of building energy performance by robust regression. Energy and Buildings, 38, pp. 582-587.

[47] Ansari, F.A., Mokhtar, A.S., Abbas, K.A., Adam, N.M., 2005. A simple approach for building cooling load estimation. American Journal of Environmental Sciences, 1, pp. 209-212.

[48] Jimenez, M.J., Heras, M.R., 2005. Application of multi-output ARX models for estimation of the U and g values of building components in outdoor testing. Solar Energy, 79, pp. 302-310.

[49] Kimbara, A., Kurosu, S., Endo, R., Kamimura, K., Matsuba, T., Yamada, A., 1995. On-line prediction for load profile of an air-conditioning system. ASHRAE Transactions, 101, pp. 198-207.

[50] Naveros, I., Ghiaus, C., Ruiz, D.P. and Castano, S., 2015. Physical parameters identification of walls using ARX models obtained by deduction. Energy and Buildings, 108, pp.317-329.

[51] Cole, W.J., Rhodes, J.D., Gorman, W., Perez, K.X., Webber, M.E. and Edgar, T.F., 2014. Community-scale residential air conditioning control for effective grid management. Applied Energy, 130, pp.428-436.

[52] Edwards, R.E., New, J., Parker, L.E., Cui, B. and Dong, J., 2017. Constructing large scale surrogate models from big data and artificial intelligence. Applied Energy, 202, pp.685-699.

[53] Dong, B., Cao, C., Lee, S.E., 2005. Applying support vector machines to predict building energy consumption in tropical region. Energy and Buildings, 37, pp. 545-553.

[54] Chen, L., Basu, B. and McCabe, D., 2016. Fractional order models for system identification of thermal dynamics of buildings. Energy and Buildings, 133, pp.381-388.

[55] Hou, Z., Lian, Z., Yao, Y., Yuan, X., 2006. Cooling-load prediction by the combination of rough set theory and an artificial neural-network based on data-fusion technique. Applied Energy, 83, pp. 1033-1046.

[56] Zhang, Y., O'Neill, Z., Dong, B. and Augenbroe, G., 2015. Comparisons of inverse modeling approaches for predicting building energy performance. Building and Environment, 86, pp.177-190.

[57] Fan, C., Xiao, F., Zhao, Y., 2017. A short-term building cooling load prediction method using deep learning algorithms. Applied Energy, 195, pp. 222-233.

[58] Fan, C., Xiao, F., Wang, S.W., 2014. Development of prediction models for next-day building energy consumption and peak power demand using data mining techniques. Applied Energy, 127, pp. 1-10.

[59] Edwards, R.E., New, J. and Parker, L.E., 2012. Predicting future hourly residential electrical consumption: A machine learning case study. Energy and Buildings, 49, pp.591-603.

[60] International Energy Conservation Codes (IECC) – 2006 IECC Prescriptive Requirements- Available at: https://www.energycodes.gov/sites/default/files/documents/ta_2006_iecc_prescriptive_requirements.pdf

[61] RESNET. Understanding the HERS® Index. Accessed 4/26/2017. Available at :http://www.hersindex.com/understanding.

[62] Christian, J. E., A.C. Gehl, P.R. Boudreaux, J. R. New. Campbell Creek TVA 2010 First Year Performance Report. Oak Ridge National Laboratory. 2010 August. No.: ORNL/TM-2010/206.

[63] Hendron, R. and Engebrecht, C. Building America house simulation protocols. Golden, CO: National Renewable Energy Laboratory; 2010.

[64] Yang SM, Zhang ZZ. An experimental study of natural convection heat transfer from a horizontal cylinder in high Rayleigh number laminar and turbulent regions. In: Hewitt G F. ed. Proceedings of the 10th International Heat Transfer Conference. Brighton, 1994, 7: 185-189.

[65] Yang SM, Jiang CJ. Criterion of transition to transitional correlation of natural convection heat transfer from horizontal cylinder in air. In: Wang BX. Ed. Heat transfer science and technology. Beijing: Higher Education Press, 1996. 181-186.

[66] Incropera FP, DeWitt DP. Introduction to heat transfer. 3rd ed. New York: John Wiley & Sons, 1996: 403-406.

[67] Holman JP. Heat transfer. 8th ed. New York: McGraw – Hill Companies, 1997. 289, 305, 364.

[68] Chen, Q., 2009. Ventilation performance prediction for buildings: A method overview and recent applications. Building and environment, 44(4), pp.848-858.

[69] Costa VAF. Natural convection in partially divided square enclosures: Effects of thermal boundary conditions and thermal conductivity of the partitions. International Journal of Heat and Mass Transfer

[70] Jin M, Liu W, Chen Q. Simulating buoyancy-driven airflow in buildings by coarse-grid fast fluid dynamics. Building and Environment 2015; 144-152.

[71] Xu, X.H. and Wang, S.W. A simplified dynamic model for existing buildings using CTF and thermal network models. International Journal of Thermal Sciences 2008; 47(9): 1249-1262.

[72] National Oceanic and Atmospheric Administration (NOAA). Solar Calculation Details – Available at: https://www.esrl.noaa.gov/gmd/grad/solcalc/calcdetails.html.

[73] ASHRAE. ASHRAE Handbook: Fundamentals (SI Edition). Atlanta, GA: American Society of Heating, Refrigerating and Air-conditioning Engineers; 2013.

[74] Domínguez-Muñoz, F., Cejudo-López, J.M. and Carrillo-Andrés, A. Uncertainty in peak cooling load calculations. Energy and Buildings 2010; 42(7): 1010-1018.

[75] Rstudio-Open source and enterprise-ready professional software for R. Available at: https://www.rstudio.com/

[76] Chen, T.Q., Guestrin, C., 2016. XGBoost: a scalable tree boosting system. KDD, San Francisco, CA, USA.

[77] Ke G.L. 2017. LightGBM: Light gradient boosting machine. R package version 0.2, URL <https://github.com/Microsoft/LightGBM>.