# Physical-data Fusion Modeling Method for Energy Consumption Analysis of Smart Building

Xiao Han, Chaohai Zhang, Yi Tang, and Yujian Ye

*Abstract*—**The energy consumption of buildings accounts for approximately 40% of total energy consumption. An accurate energy consumption analysis of buildings can not only promise significant energy savings but also help estimate the demand response potential more accurately, and consequently brings benefits to the upstream power grid. This paper proposes a novel physical-data fusion modeling (PFM) method for modeling smart buildings that can accurately assess energy consumption. First, a thermal process model of buildings and an electrical load model that focus on building heating, ventilation, and air conditioning (HVAC) systems are presented to analyze the thermal-electrical conversion process of energy consumption of buildings. Second, the PFM method is used to improve the accuracy of the energy consumption analysis model for buildings by modifying the parameters that are difficult to measure in the physical model (i. e., it effectively modifies the electrical load model based on the proposed PFM method). Finally, case studies involving a real-world dataset recorded in a high-tech park in Changzhou, China, demonstrate that the proposed method exhibits superior performance with respect to the traditional physical modeling (TPM) method and data-driven modeling (DDM) method in terms of the achieved accuracy.**

*Index Terms*—**Smart building, physical-data fusion modeling method, energy consumption, precision model, thermal-electrical conversion.**

## I. Introduction

WITH the continuous increase in urbanization and industrial restructuring, the energy consumption of buildings in urban cities has increased rapidly over the world [1], [2]. According to [3], the energy consumption of buildings accounts for 40% of total energy consumption, and approximately 80% of energy consumption of buildings is attributed to heating, ventilation, and air conditioning (HVAC) systems [3], [4]. Therefore, driven by the rapid development and deployment of smart sensors, information, communication, and artificial intelligence (AI) technologies [5], the energy consumption of smart buildings can be analyzed and evaluated more accurately to achieve energy conservation [6], [7]. As a result, it promises a significant demand response potential to support the secure and economic operation of power grids [8], [9].

With the advances in physical measurement tools and monitoring systems for buildings, the energy consumption can be measured more accurately [10]. Modeling methods for energy consumption of buildings can be classified into three types, i. e., physical modeling methods, data-driven modeling (DDM) methods, and physical-data fusion modeling (PFM) methods [11].

Physical modeling methods can be further subdivided into simplified and precise physical models [12]. Simplified physical models make certain assumptions and simplify the complex physical processes in building operations, such as the interaction of internal areas, thermal processes, and air convection [13]. These modeling methods also neglect many physical quantities that are difficult to be measured such as the heat exchange coefficient and energy efficiency ratio (EER) of HVAC systems. In addition, the simplified physical models are not very scalable. Therefore, they pose considerable challenges in accurately accounting for the time-varying thermal-electrical conversion process of a building, which leads to poor generalization performances.

Precise modeling methods construct the models for thermal processes, fluids, and electrical equipment of buildings and then combine them to build complex physical process models. Therefore, a precise physical model is generally implemented using available commercial software such as eQUEST, PowerDOE, DeST, and FloVENT [14], which partially consider certain complexities associated with the physical models of buildings. However, these software programs have certain limitations.

1) The time-varying characteristics of building information at multiple time scales are not properly considered in these programs [15].

2) The energy consumption analysis of buildings may require the use and coordination of multiple software programs, which are assigned different tasks according to their design features.

Reference [16] uses eQUEST to simulate the energy consumption of buildings and then uses FloVENT to simulate the heat distribution in the internal area of buildings. Both software programs are limited in their abilities to consider

Manuscript received: January 23, 2021; revised: June 1, 2021; accepted: November 3, 2021. Date of Cross Check: November 3, 2021. Date of online publication: February 4, 2022.

This work was supported in part by the National Natural Science Foundation of China (No. 51877037).



X. Han and C. Zhang (corresponding author) are with the Jiangsu Key Laboratory of New Energy Generation and Power Conversion, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China (e-mail: hanxiao0625@163.com; zhangchaohai@nuaa.edu.cn).

Y. Tang is with the School of Electrical Engineering, Southeast University, Nanjing 210096, China (e-mail: tangyi@seu.edu.cn).

Y. Ye is with the Department of Electrical and Electronic Engineering, Imperial College London, London, U.K. (e-mail: yujian.ye11@imperial.ac.uk).

DOI: 10.35833/MPCE.2021.000050

the effects of critical aspects such as the social and behavioral factors of building residents on building modeling tasks.

DDM or statistical modeling methods are based on various statistical or machine learning methods to describe or predict the energy consumption of buildings through extensive real-time and historical measurement data. These methods do not rely on the accurate modeling of physical characteristics of buildings [17], [18]. The data collection is usually implemented through real-time online monitoring systems such as the building energy management system (BEMS) [19]. Through the collection of a large amount of environmental and electrical data by the BEMS, DDM methods can achieve relatively high accuracy through extensive training on large-scale datasets [20]. However, the training accuracy of DDM methods can be significantly affected by the poor data quality of the samples [2], [4].

The results of DDM methods mostly depend on the quality of historical data. However, historical data have many problems such as poor data quality and low data density, which result in a loss of information. Consequently, the support from sophisticated algorithms is required [21]. In [21], an energy consumption modeling method is proposed to solve the problem of lacking the information about buildings, and the accuracy of this method is verified through its application to multiple buildings. These buildings may have the equipment that is missing or damaged or have communication problems. Simultaneously, other cases may occur in special scenarios such as with equipment maintenance and during holidays, which may lead to not only missing data but also incorrect data. In [22], a modeling method of data relevance is considered, and the comprehensiveness of the model is improved by integrating external information such as the external environment, social policies, and user behaviors. However, this method has poor physical interpretation and is difficult to understand. In addition, there may be different situations of incomplete information collection in different environments, which is not considered in the DDM methods for buildings. Furthermore, the accuracy of measurement data significantly affects the accuracy of the model. Moreover, the calculation algorithm plays a key role in improving the accuracy of the model. In [17], a bicubic interpolation algorithm and convolutional neural network are used for the spatial prediction of energy consumption. In [18], the evolutionary algorithms are discussed to solve the model for energy consumption of buildings. However, a heavy computational burden makes it difficult to apply AI algorithms to online analysis on the energy consumption of buildings. In other words, the simplified online DDM methods neglect the physical characteristics and causality of buildings.

Therefore, the main contribution of this study is to establish a detailed physical model for smart buildings and ensure that the energy consumption analysis is more precise by modifying the model through a PFM method. The energy consumption analysis model of buildings is based on the PFM method, which considers not only the interpretation ability of the physical mechanism but also the incidence relation of DDM methods. Physical methods can provide high-entropy information for DDM methods. In addition, physical methods can ensure that the parameters of physical model are more targeted and can prevent the modeling method from falling into a partial learning space. The critical factor is that an accurate analysis model for energy consumption of buildings based on the PFM method does not require extensive training data and can be run online. Finally, the PFM method has a high initial training accuracy rate, which means that this method is more scalable. Therefore, a detailed and accurate physical model needs to be constructed using DDM methods to compensate for the loss of rules derived from the simplification of physical processes.

The remainder of this paper is organized as follows. Section II presents the thermal process model that considers the structures and internal thermal-electrical conversion processes of buildings. Section III investigates a solution for improving the accuracy of the model, namely, a PFM method, which is used first to modify the key parameters in the physical model and then to determine the accuracy of the power load model while also providing a replacement strategy. The results of case studies are presented in Section IV, and conclusions are given in Section V.

## II. Precise Physical Models of Buildings

### A. Dataset Collection of Buildings

The premise of establishing a precise energy consumption analysis model for buildings is to create a comprehensive and high-quality dataset. The dataset should include structural, environmental, electrical, and behavioral data of buildings, as shown in Fig. 1.

In addition, with a large amount of heterogeneous data, it is necessary to consider multiple time-scale characteristics and perform data preprocessing.

The structural data of buildings mainly include measurement parameters such as area and floor height, structural parameters such as orientation and shape, material parameters such as roof materials, wall materials, and vent types, and empirical parameters such as heat transfer coefficients and specific heat capacities. Because most of this information is static data, the data processing is not necessary. However, problems of inaccurate measurement data and empirical parameters often arise and affect the accuracy of physical model. Therefore, a PFM method needs to be employed to solve these problems.

Correspondingly, the environmental information, electrical data, and behavioral data are dynamic parameters. These datasets can be collected by monitoring equipment such as circuit breakers and smart plugs, intelligent sensors, and environmental monitoring devices installed both inside and outside the building. The environmental data are categorized into external and internal environmental data. These data mainly include temperature, humidity, light intensity, and climate data, etc. The acquisition of these data can also be divided into fixed-period collection and trigger-mode collection. The trigger-mode collection means that some data such as $CO_2$ concentration do not change considerably over a short period. We only needs to wait until a significant change is detected before collecting the data. Electrical data include the voltage, current, frequency, active power, reactive power, and accumulated electrical energy collected by each electrical device and each power node, EER, and coefficient of performance (COP), personal computer (PC) load, photovoltaic (PV), etc. Based on different types of electrical equipment, the collection period can be set to be seconds, minutes, or hours. Finally, the behavioral data include status data and operational logs of electrical equipment, doors, windows, and other vents as well as personnel data for each area, running scenarios, and others. Most of these data are collected in trigger mode.

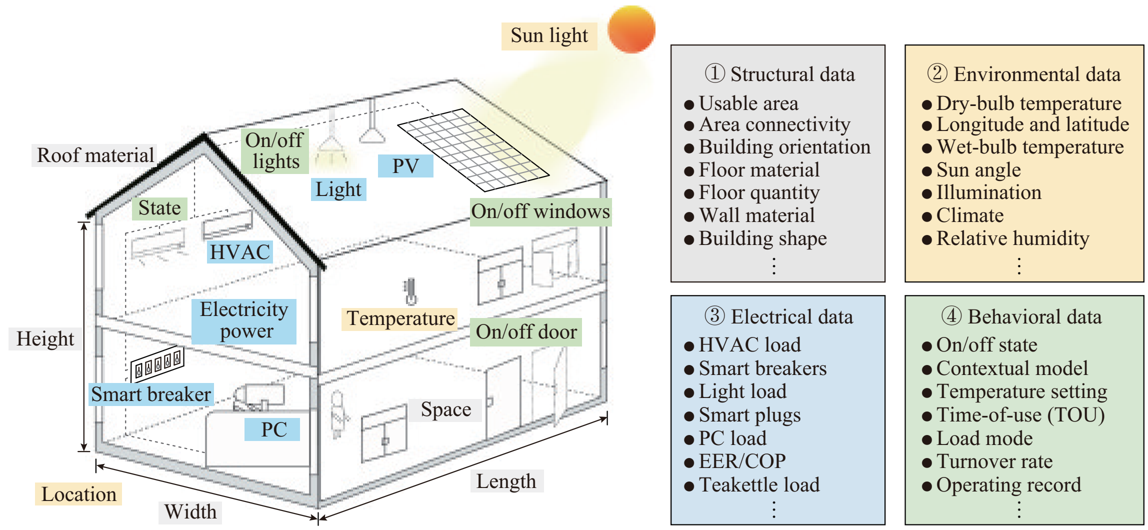


Fig. 1. Dataset collection system of building.

### B. Structural Model of Buildings

To establish a refined physical model of a building, it is necessary to perform a detailed model of several internal areas such as rooms and aisles in the entire building. In this manner, the effects of the external environment and internal areas of the building can be fully considered. Simultaneously, this can also improve the accuracy of the thermal process and energy consumption analysis. Therefore, this study constructs two matrices to describe the structural relationships and corresponding parameters of each internal area of a building. The first is the structural association matrix of the building, which is given as:

$$\boldsymbol{A}_{nect}^{(m)}=\begin{bmatrix} 0 & b_{ven,01} & \cdots & b_{ven,0n} \\ b_{con,10} & a_1 & \cdots & b_{ven,1n} \\ \vdots & \vdots & & \vdots \\ b_{con,n0} & b_{con,n1} & \cdots & a_n \end{bmatrix} \tag{1}$$

where $\boldsymbol{A}_{nect}^{(m)}$ is the internal structural association matrix of the $m^{\text{th}}$ building; $a_i$ is the property of the $i^{\text{th}}$ area, which includes rooms, open areas, corridors, etc; $b_{con,ij}$ denotes whether one or more enclosure structures exist between the $i^{\text{th}}$ and $j^{\text{th}}$ areas of the building; $b_{ven,ji}$ denotes whether one or more vent structures exist between the $j^{\text{th}}$ and $i^{\text{th}}$ areas of the building; and $n$ is the number of areas in the building.

Next, the structural parameter matrix of the building is given as (2), which provides detailed information about the internal areas.

$$\boldsymbol{A}_{parm}^{(m)}=\begin{bmatrix} \boldsymbol{0} & \boldsymbol{d}_{ven,01} & \cdots & \boldsymbol{d}_{ven,0n} \\ \boldsymbol{d}_{con,10} & \boldsymbol{c}_1 & \cdots & \boldsymbol{d}_{ven,1n} \\ \vdots & \vdots & & \vdots \\ \boldsymbol{d}_{con,n0} & \boldsymbol{d}_{con,n1} & \cdots & \boldsymbol{c}_n \end{bmatrix} \tag{2}$$

where $\boldsymbol{A}_{parm}^{(m)}$ is the structural parameter matrix of the $m^{\text{th}}$ building; $\boldsymbol{c}_i$ contains the volume of the $i^{\text{th}}$ area inside the building; $\boldsymbol{d}_{con,ij}$ and $\boldsymbol{d}_{ven,ji}$ are the parameter matrices of the enclosure structures and vents between the $i^{\text{th}}$ and $j^{\text{th}}$ areas, respectively, as given in (3) and (4).

$$\boldsymbol{d}_{con,ij}=\begin{bmatrix} s_{con,ij}^{1} & h_{con,ij}^{1} & o_{con,ij}^{1} & r_{con,ij}^{1} \\ s_{con,ij}^{2} & h_{con,ij}^{2} & o_{con,ij}^{2} & r_{con,ij}^{2} \\ \vdots & \vdots & \vdots & \vdots \\ s_{con,ij}^{vc} & h_{con,ij}^{vc} & o_{con,ij}^{vc} & r_{con,ij}^{vc} \end{bmatrix} \tag{3}$$

$$\boldsymbol{d}_{ven,ji}=\begin{bmatrix} s_{ven,ji}^{1} & h_{ven,ji}^{1} & o_{ven,ji}^{1} & r_{ven,ji}^{1} \\ s_{ven,ji}^{2} & h_{ven,ji}^{2} & o_{ven,ji}^{2} & r_{ven,ji}^{2} \\ \vdots & \vdots & \vdots & \vdots \\ s_{ven,ji}^{vv} & h_{ven,ji}^{vv} & o_{ven,ji}^{vv} & r_{ven,ji}^{vv} \end{bmatrix} \tag{4}$$

where the superscripts $vc$ and $vv$ are the numbers of envelope structures and vents between the $i^{\text{th}}$ and $j^{\text{th}}$ areas of the building, respectively; $s_{con,ij}$, $h_{con,ij}$, $o_{con,ij}$, and $r_{con,ij}$ are the acreage, heat transfer coefficient, material type, and orientation of the envelope structures between the $i^{\text{th}}$ and $j^{\text{th}}$ areas of the building, respectively; and $s_{ven,ji}$, $h_{ven,ji}$, $o_{ven,ji}$, and $r_{ven,ji}$ are the acreage, heat transfer coefficient, material type, and orientation of the vents between the $i^{\text{th}}$ and $j^{\text{th}}$ areas of the building, respectively.

### C. Thermal Process Model of Buildings

A dynamic thermal process model of a building needs to be established based on the structural model of building. Taking a single area inside the building as an example, the dynamic thermal process of a single area refers to the heat transfer process of internal and external disturbances, as shown in Fig. 2. The external heat transfer $q_{EXT}$ includes the outdoor environment (e.g., ambient temperature and solar radiation) and external thermal conditions (e.g., thermal conditions in adjacent areas). The external heat transfer mainly affects the internal temperature of the room through the heat transfer through the wall and other enclosure structures or through windows and other vents. Indoor air is affected by heat convection and heat radiation. By contrast, internal heat transfer $q_{INT}$ includes latent heat and sensible heat of furniture, equipment, and human bodies. The latent heat of human bodies and equipment directly acts on the air in an area, which can immediately devalue the air in the area. The sensible heat of HVAC, lights, computers, and other equipment is directly transferred to indoor air through thermal convection.

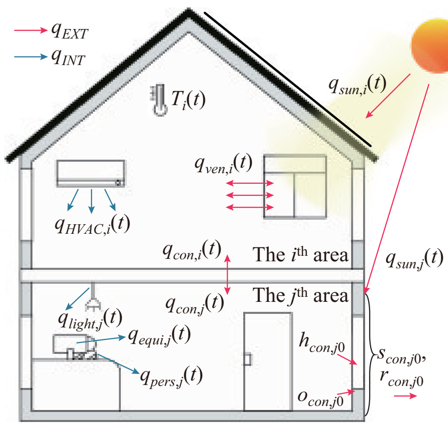


Fig. 2. Dynamic thermal process model for a building.

To obtain a refined thermal process model for each area in a building, it is important to solve the partial differential equations for heat transfer of the enclosure structure in each area. In addition, the connectivity between regions needs to be considered. The model of dynamic thermal process for a single area is given as:

$$\frac{\mathrm{d}T_i(t)}{\mathrm{d}t} = \frac{1}{C_i \rho_i c_i}(q_{EXT,i}(t) + q_{INT,i}(t)) \tag{5}$$

where $T_i(t)$ is the temperature of the $i^{\text{th}}$ area inside the building at time $t$; $\mathrm{d}T_i(t)/\mathrm{d}t$ is the change of temperature in the $i^{\text{th}}$ area inside the building at time $t$; $C_i$ is the specific heat capacity of air in the $i^{\text{th}}$ area; $\rho_i$ is the air density in the $i^{\text{th}}$ area; $q_{EXT,i}(t)$ is the heat transferred by the external disturbance heat process in the $i^{\text{th}}$ area at time $t$; and $q_{INT,i}(t)$ is the heat transferred by the internal disturbance heat process in the $i^{\text{th}}$ area at time $t$.

Based on (5) and Fig. 2, the heat obtained by the external disturbance can be further calculated using (6) - (9). Heat transferred by the external disturbance heat process in buildings mainly includes heat transfer from the enclosure structure, vents, and sun radiation.

$$q_{EXT,i}(t) = q_{con,i}(t) + q_{ven,i}(t) + q_{sun,i}(t) \tag{6}$$

$$q_{con,i}(t) = \begin{cases} \sum\limits_{j=0}^{n} \sum\limits_{k=1}^{A_{nect}(j,i)} \Delta T(t) \prod\limits_{\mu=1}^{2} d_{con,ji}(k,\mu) & i<j \\ \sum\limits_{j=i+1}^{n} \sum\limits_{k=1}^{A_{nect}(i,j)} \Delta T(t) \prod\limits_{\mu=1}^{2} d_{con,ij}(k,\mu) & i>j \end{cases} \tag{7}$$

$$\Delta T(t) = T_j(t) - T_i(t) \tag{8}$$

$$q_{ven,i}(t) = \begin{cases} \sum\limits_{j=0}^{n} \sum\limits_{k=1}^{A_{nect(j,i)}} \Delta T(t) \prod\limits_{\mu=1}^{2} d_{ven,ji}(k,\mu) & j<i \\ \sum\limits_{j=i+1}^{n} \sum\limits_{k=1}^{A_{nect(i,j)}} \Delta T(t) \prod\limits_{\mu=1}^{2} d_{ven,ij}(k,\mu) & j>i \end{cases} \tag{9}$$

where $q_{con,i}(t)$, $q_{ven,i}(t)$, and $q_{sun,i}(t)$ are the heat transferred from the external area to the $i^{\text{th}}$ area through the enclosure structure, vent, and solar radiation at time $t$, respectively, and $q_{sun,i}(t)$ is calculated by measuring the intensity of light [23], [24]; $\Delta T(t)$ is the temperature difference between the $i^{\text{th}}$ and $j^{\text{th}}$ areas; and $A_{nect(i,j)}$ and $A_{nect(j,i)}$ are the numbers of enclosure structures and vents between the $i^{\text{th}}$ and $j^{\text{th}}$ areas inside the building, respectively.

The heat transferred by the internal disturbance heat process in buildings mainly includes heat transfer from HVAC, lights, personnel, and other equipment, as shown in (10).

$$q_{INT,i}(t) = q_{HVAC,i}(t) + q_{light,i}(t) + q_{pers,i}(t) + q_{equi,i}(t) \tag{10}$$

where $q_{HVAC,i}(t)$, $q_{light,i}(t)$, $q_{pers,i}(t)$, and $q_{equi,i}(t)$ are the heat transferred from the HVAC, lights, personnel [23], and other equipment to the $i^{\text{th}}$ area at time $t$, respectively, and $q_{HVAC,i}(t)$, $q_{light,i}(t)$, and $q_{equi,i}(t)$ are calculated by measuring or predicting the power of the corresponding equipment.

### *D. Thermal-electrical Conversion Model of Buildings*

To analyze the energy consumption of buildings, a thermal-electrical conversion model of buildings is established. First, the total power calculation model for a building is given as:

$$P(t) = \sum_{i=0}^{n} (P_{HVAC,i}(t) + P_{light,i}(t) + P_{equi,i}(t)) - P_{PV,i}(t) \tag{11}$$

$$\begin{cases} P_{HVAC,i}(t) = g_{HVAC}(K_{NPLV,i}, x_{ws,i}, \eta_i) q_{HVAC,i} \\ P_{light,i}(t) = g_{light}(K_{light,i}, x_{stlight,i}) q_{light,i} \\ P_{equi,i}(t) = g_{equi}(K_{equi,i}, x_{stequi,i}) q_{equi,i} \end{cases} \tag{12}$$

where $P_{HVAC,i}(t)$, $P_{light,i}(t)$, $P_{equi,i}(t)$, and $P_{PV,i}(t)$ are the electric power of HVAC, lights, other equipment, and PV generation in the $i^{\text{th}}$ area at time $t$; $g_{HVAC}(\cdot)$, $g_{light}(\cdot)$, and $g_{equi}(\cdot)$ are the thermal-electric conversion coefficient functions for HVAC, lights, and other equipment, respectively; $K_{NPLV,i}$, $K_{light,i}$, and $K_{equi,i}$ are the thermal-electric conversion coefficients of HVAC, lights, and other equipment in the $i^{\text{th}}$ area, respectively; $x_{ws,i}$, $x_{stlight,i}$, and $x_{stequi,i}$ are the operating states of the HVAC, lights, and other equipment in the $i^{\text{th}}$ area, respectively; and $\eta_i$ is the thermal-electric conversion efficiency.

$$x_{ws} = \{x_{mode}, x_{st}, x_{speed}\} \tag{13}$$

$$g_{HVAC}(K_{NPLV}, x_{ws}, \eta) = K_{NPLV}(x_{mode})(1 + \eta(x_{speed})) \tag{14}$$

where $x_{mode}$ is the operating mode of HVAC (refrigeration, heating, air supply); $x_{st}$ is the operating switch status (on/off) of HVAC; $x_{speed}$ is the wind speed of HVAC; and $K_{NPLV}$, $K_{light}$, and $K_{equi}$ are the EER corresponding to HVAC, lights, and other equipment, respectively, which indicates the inherent energy efficiency of equipment. Under the specific circumstances, the EERs of different devices are differently defined. The calculation methods for the energy consumption and power of other electrical loads (e.g., lights, water boilers, PCs) in buildings are similar, and the formulas are shown as:

$$I(\Gamma)=\sum_{i=0}^{n}(I_{HVAC,i}(\Gamma)+I_{light,i}(\Gamma)+I_{equi,i}(\Gamma))-I_{PV,i}(\Gamma) \tag{15}$$

$$\begin{cases} I_{HVAC,i}(\Gamma)=\int_{\Gamma}P_{HVAC,i}(t)\mathrm{d}t \\ I_{light,i}(\Gamma)=\int_{\Gamma}P_{light,i}(t)\mathrm{d}t \\ I_{equi,i}(\Gamma)=\int_{\Gamma}P_{equi,i}(t)\mathrm{d}t \end{cases} \tag{16}$$

where $\Gamma$ is the period required to calculate the energy consumption; and $I_{HVAC,i}(\Gamma)$, $I_{light,i}(\Gamma)$, $I_{equi,i}(\Gamma)$, and $I_{PV,i}(\Gamma)$ are the energy consumptions of HVAC, lights, other equipment, and PV generation in the $i^{\text{th}}$ area during time period $\Gamma$, respectively.

## III. PFM Method

### A. PFM Algorithm

The basis of the PFM method is the fusion mode. The physical analysis method can provide high-entropy information, which helps improve the efficiency of the data model analysis. In other words, the input features contain the target features to be predicted, which can narrow the search space and reduce the computational complexity when solving the parameters of data model through the optimization process. However, the PFM method can be used to construct a better data model with high-entropy input features, achieving the goal of building a data model. This also ensures that the model parameter optimization is more targeted. To avoid falling into local optimization, the PFM method improves the rationality of the data model. This approach can compensate for the loss of physical discipline rules in physical analysis methods due to the model simplification.

This study proposes two PFM correction methods. One (method 1) modifies the key parameters in the physical model through the DDM method; the other (method 2) replaces the submodels in the PFM method through the model modification. The process flow of method 1 is shown in Fig. 3.

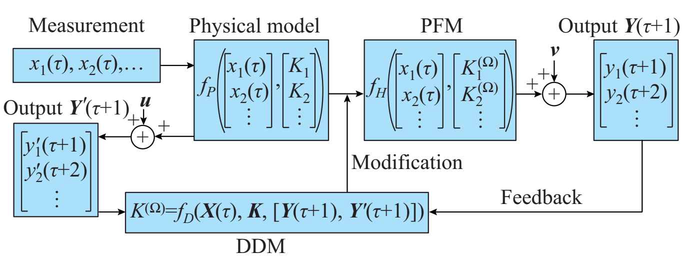


Fig. 3. Process flow of method 1.

The PFM method for energy consumption analysis of smart buildings requires the configuration of model parameters. The parameters include the static parameters that are input when the model is built and the dynamic parameters obtained in real-time collection when the model is running. In this paper, the parameters of thermal-electrical conversion model for a smart building that need to be modified mainly include the heat exchange coefficient and EER. Therefore, the final PFM can be obtained by guiding and correcting a simple physical model using the measurement data. The methodology for correcting the key parameters in the physical model through a DDM method is given as:

$$\boldsymbol{Y}'(\tau+1)=f_P(\boldsymbol{X}(\tau),\boldsymbol{K})+\boldsymbol{u}=f_H(\boldsymbol{X}(\tau),\boldsymbol{K}^{(\Omega)})+\boldsymbol{v} \tag{17}$$

$$\boldsymbol{K}^{(\Omega)}=f_D(\boldsymbol{X}(\tau),\boldsymbol{K},[\boldsymbol{Y}(\tau+1),\boldsymbol{Y}'(\tau+1)]) \tag{18}$$

where $f_P$ and $f_D$ are the functions of physical model and DDM methods, respectively; $f_H$ is the function of PFM method, which uses the same algorithm as $f_P$; $\boldsymbol{X}(\tau)$ is the input dataset of the model at time $\tau$; $\boldsymbol{Y}(\tau+1)$ and $\boldsymbol{Y}'(\tau+1)$ are the output datasets of the model at time $\tau+1$; $\boldsymbol{K}$ is the parameter vector in the model; $\boldsymbol{K}^{(\Omega)}$ is the parameter vector after DDM correction; and $\boldsymbol{u}$ and $\boldsymbol{v}$ are the random errors. In this study, $\boldsymbol{K}^{(\Omega)}$ may include the parameters such as the heat exchange coefficient and EER.

Figure 4 depicts the process flow for modifying the thermal-electrical conversion model of the building using the PFM (method 2). The physical and statistical mapping models are represented by (19) and (20), respectively.

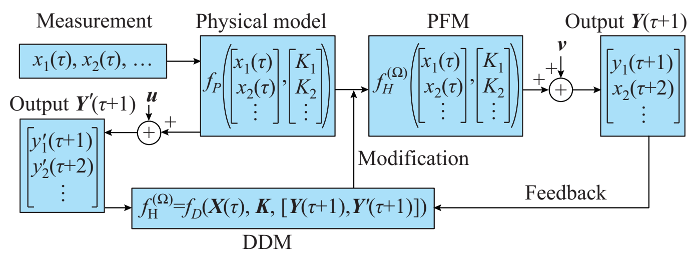


Fig. 4. Process flow of method 2.

$$\boldsymbol{Y}'(\tau+1)=f_P(\boldsymbol{X}(\tau),\boldsymbol{K})+\boldsymbol{u}=f_H^{(\Omega)}(\boldsymbol{X}(\tau),\boldsymbol{K})+\boldsymbol{v} \tag{19}$$

$$f_H^{(\Omega)}=f_D(\boldsymbol{X}(\tau),\boldsymbol{K},[\boldsymbol{Y}(\tau+1),\boldsymbol{Y}'(\tau+1)]) \tag{20}$$

where $f_H^{(\Omega)}$ is the function set of the modified PFM, which is obtained after replacing the submodels with DDM. The replacement method and selection mechanism of submodels in PFM are later analyzed in detail in conjunction with the building.

### B. Key Parameter Modification Based on PFM

Based on the PFM method, this study divides the process for energy consumption analysis of the building into two steps. The first step involves modifying the key parameters in the precise physical model through DDM method. The second step involves modifying some of the submodels in the entire energy consumption analysis model and then building a model selection mechanism to improve the accuracy of PFM. The overall process is shown in Fig. 5.

In Section II, a precise physical model of the building is established. Next, the key parameter modification in a physical model using DDM method is discussed. The parameters to be modified include the heat transfer coefficient of enclosure structures and vents and the conversion efficiency of HVAC. These parameters are difficult to be measured or calculated. For example, the heat transfer coefficient can be affected by the wall material, shape, aging degree, and other factors.

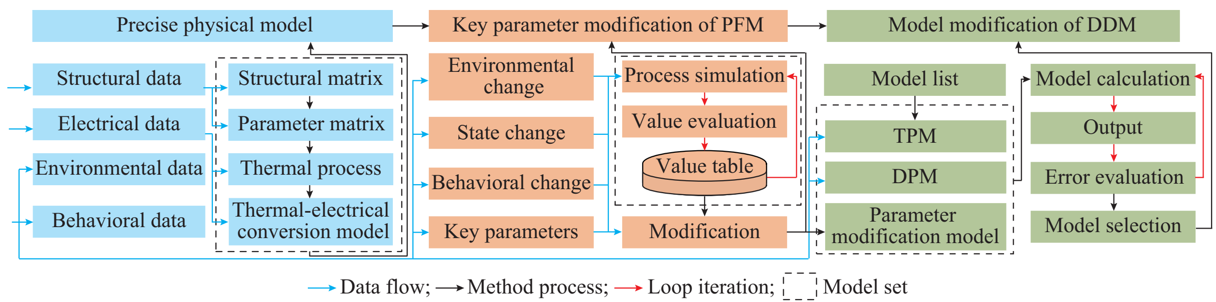


Fig. 5. PFM process for energy consumption analysis of building.

In this paper, a long short-term memory (LSTM) algorithm is used to modify the key parameters. LSTM has been a popular machine learning algorithm in recent years [19]. Because of the temporal characteristics of the electrical, environmental, and behavioral data of buildings, LSTM is selected because it has advantages in dealing with time-sequenced modeling. In addition, LSTM can be used to process global sample datasets.

As a recurrent neural network, the input, output, and error calculation forms of LSTM are explained in this study. The input of the DDM method is a matrix. The horizontal axis represents the characteristic quantity of sample datasets that includes electrical, environmental, and behavioral information as well as key parameters that must be modified. The vertical axis represents the time-sequenced values. In the case studies, each sample dataset is selected for one day, with an interval period of 5 min. The output of LSTM is divided into two data parts: thermal process data of various areas in the building including the external and internal heat exchange at various time, and power and status information of different power loads in the building. The error is generated between the output of each loop and the target measurement. The error is also calculated backwards, affecting every gate in the output back to the input stage until this value is filtered out.

## C. Model Modification and Selection Mechanism

The second process involves modifying the submodels. In the process of energy consumption analysis of an entire building, problems may occur in which the physical modeling of some links is inaccurate or the parameters are difficult to modify. Therefore, this study proposes a model modification mechanism. Other modeling methods can be used to replace some of the modeling aspects in the energy analysis, such as thermal process links, HVAC modeling, or other electrical load modeling, as shown in Fig. 5.

The HVAC model is selected as a submodel for discussion. This study presents two typical HVAC algorithms and provides a simple introduction. First, the modification and selection mechanism in this study are discussed through these two typical models. Second, two models, i.e., the traditional physical modeling (TPM) and DDM are used for comparison purposes in the case studies to evaluate the effects of the PFM proposed in this paper.

### 1) TPM

In [25], a classic HVAC model is proposed to evaluate the HVAC energy consumption through three types of temperature parameters, i.e., outdoor, indoor, and set temperatures, combined with air conditioning power and cooling/heating efficiency. The model also adds the users' comfort limit on the power consumption of HVAC, as shown in (21) and (22).

$$C_{i,t}^{HVAC}=\gamma_i(\theta_{i,t-1}^{in}-\theta_{i,t}^{ref})^2 \tag{21}$$

$$\theta_{i,t}^{in}=\theta_{i,t-1}^{in}+\alpha_i(\theta_{i,t}^{out}-\theta_{i,t-1}^{in})-\beta_i q_{i,t}^{HVAC} \tag{22}$$

where $C_{i,t}^{HVAC}$ is the cost of HVAC considering the users' comfort [25]; $\theta_{i,t}^{in}$ and $\theta_{i,t}^{out}$ are the indoor and outdoor temperatures of building $i$ at time $t$, respectively; $\gamma_i$ is the environmental preferences (i.e., the temperature sensitivities) of the occupants in the building $i$; $\alpha_i$ is the thermal insulation; and $\beta_i$ is the energy efficiency of HVAC.

In [22], an HVAC model could be dynamically adjusted automatically by personnel according to their levels of comfort. The model considers the effects of air temperature, average radiant temperature, relative humidity, air speed, clothing insulation, and metabolic rates on levels of comfort.

$$T_{t+1}^{in}=\varepsilon T_t^{in}+(1-\varepsilon)\left(T_t^{out}-\frac{\eta_{HVAC}}{A}e_t\right) \tag{23}$$

where $T_t^{in}$ and $T_t^{out}$ are the indoor and outdoor temperatures of the building, respectively; $\varepsilon$, $\eta_{HVAC}$, and $A$ are the conversion coefficients, and their specific meanings and values are described in [22]; and $e_t$ is the power of HVAC at time $t$.

### 2) DDM

In [19], the LSTM is also used to predict the energy consumption of residential buildings. The proposed approach integrates preprocessing and data organization mechanisms to refine the data and remove anomalies. A deep learning network is adopted to input refined sequence data into a convolutional neural network through a multi-layer bidirectional LSTM network to learn sequence patterns effectively. Finally, a high-precision result is obtained in the evaluation of residential energy consumption of the buildings.

It should be noted that the role of the data-driven module in PFM is to modify the key parameters and to select submodels. However, the DDM mentioned above is an option in the optional model list for the HVAC submodel. Furthermore, an accuracy evaluation method needs to be established to improve the model correction and selection mechanism. This paper chooses the conventional average relative error to measure the accuracy of the algorithm fitting. The algorithm also calculates the average relative error from the two perspectives of area temperature and total electricity consumption of the building.

## IV. Case Studies

### A. Application and Data Introduction

In this case study, a high-tech park with multiple types of buildings in Changzhou, China, is selected as an example. The park covers an area of approximately 4 $km^2$ in eastern China, which locates in the geometric center of the Yangtze River Delta. It has a subtropical monsoon climate with four distinct levels. Approximately 40 buildings are located in the park, which include office buildings, dormitories, factories, and research institutes. Approximately 5000 employees work in the park. Two 10 kV cables supply the power to the entire park, and the monthly electricity consumption of this park is approximately $65\times10^6$ kWh. There are abundant forms of energy supplies in the park such as distributed PV, energy storage, and geothermal energy. Energy monitoring systems have been installed for most of the buildings in the park to achieve the data collection and the smart management and control.

In this study, a typical office building in the park is selected, which is an L-shaped structure with four floors and a total area of approximately 11000 $m^2$. The envelope structure of the building is mainly composed of a steel tube concrete frame. The interior of the building can be divided into many areas that include power distribution rooms, offices, laboratories, open office areas, corridors, and halls. The energy supply form of the building is mainly composed of the power grid and rooftop PV. The main energy load of the building consists of HVAC, office computers, lights, and water boilers. The number of daily office and maintenance personnel in the building is approximately 80. The time periods of personnel activities are from 08: 00 to 20: 00 in summer and from 08:30 to 19:00 in winter.

Both environmental energy monitoring system and BEMS are installed in the building. Environmental monitoring outside the building mainly relies on a miniature weather station installed on the roof. The weather station monitors the environmental information, including temperature, humidity, light intensity, carbon dioxide concentration, PM2.5, wind speed, and rainfall. The monitoring period is 5 min. Similarly, the environmental monitoring inside the building is realized using numerous environmental monitoring sensors throughout the building. The BEMS measures the amount of distributed PV power generation and the power consumption of various internal areas. The electricity consumption in the internal areas inside the building is mainly composed of the electricity consumption of HVAC, lights, office computers, and other equipment. The collection period of the electrical quantity is also 5 min. Simultaneously, the building is equipped with a variety of smart sensing devices to monitor the opening and closing statuses of doors, windows, and the equipment, and the flow of people throughout the building. Most of these information collection cycles are triggered. The buildings selected in this study have a year-and-a-half's worth of sample data.

### B. Performance Comparison of Three Modeling Methods

Typical days in summer and winter are selected as the calculation periods. Based on the two aspects of regional temperature and energy consumption of the building, the accuracies of the PFM, TPM, and DDM methods are compared. First, a refined physical model is established for the building selected in the calculation example to analyze the energy consumption. The LSTM algorithm is then used to modify the key parameters in the refined physical model to obtain the temperature and electric power of each area in the building. The partial load model in the PFM is replaced with a TPM method to obtain the TPM-based calculation results. The partial load model in the PFM is replaced with a DDM, and the DDM-based calculation results are obtained. Finally, the results obtained by PFM and other two methods on a typical day in summer are compared, as shown in Fig. 6.

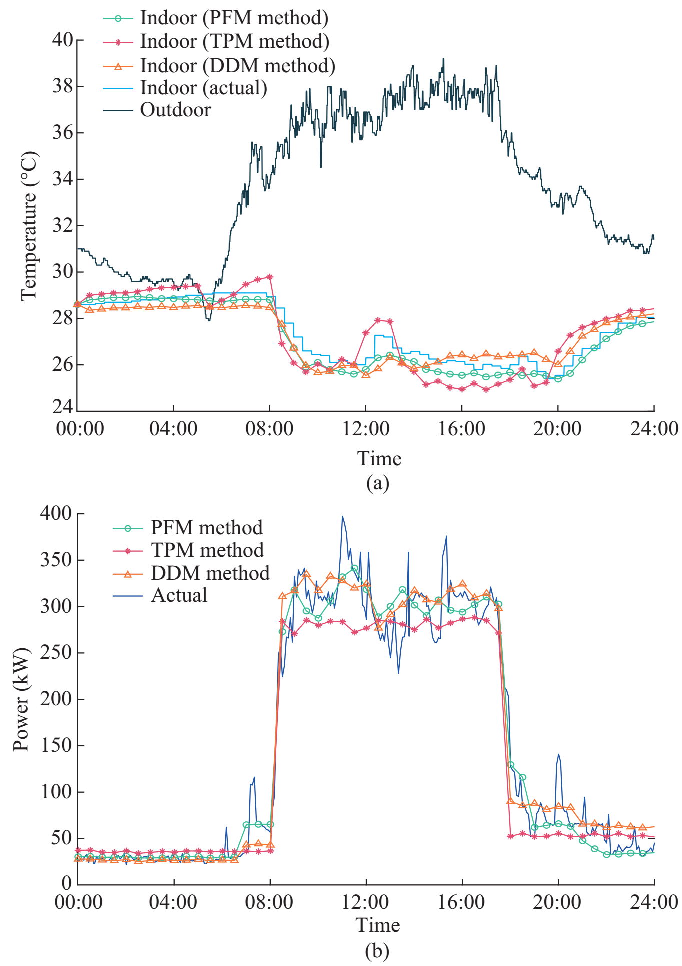


Fig. 6. Results of comparison between PFM and other two methods on a typical day in summer. (a) Comparison of regional temperature. (b) Comparison of regional energy consumption.

It can be observed from Fig. 6 that the temperature and power obtained based on the PFM method are more in line with the actual values. Therefore, the accuracy rate is higher than that of the TPM and DDM methods.

Figure 7 presents the box plot of accuracy comparison of the PFM and other two methods on a typical day in summer. The box is divided into three horizontal lines, which are the upper quartile, median, and lower quartile of accuracy, respectively. The edges of the two vertical lines on the top and bottom of the box indicate the top and bottom edges of accuracy, respectively. Some discrete points can also be observed, which represent some outliers that occasionally appear. In the calculation example, 100 sets of sample data are used to verify the accuracy results. The accuracy of PFM method is approximately 91.3% and clearly higher than that of the other two methods, and it is relatively stable (revealed by the fact that the height of the cabinet is relatively low). The TPM method has the lowest accuracy and the worst scalability. The DDM method has the medium calculation accuracy and scalability with a small amount (100 to 300 days) of sample data.

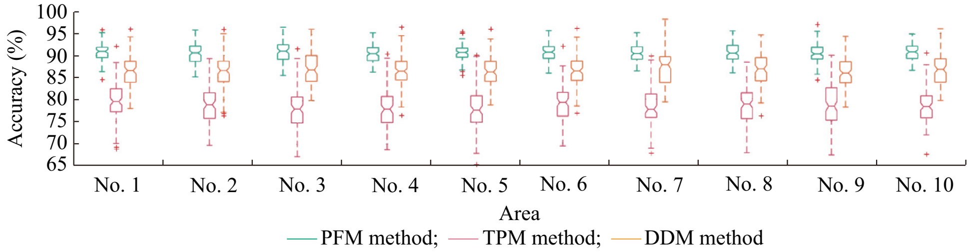


Fig. 7. Box plot of accuracy comparison between PFM and other two methods on a typical day in summer.

We then consider an example of a typical day in winter. Three differences from the calculation examples on a typical day in summer are observed. First, the working hours of employees in winter are shorter than those in summer; therefore, it is necessary to fully consider the behavioral parameters of employees in the building. Second, the working mode of HVAC changes from cooling to heating; therefore, the operating status of HVAC varies, which makes HVAC have different electrical characteristics. The third point is that in summer, the main function of HVAC is cooling, but the solar radiation increases the temperature, which is the opposite of air conditioning. However, in winter, the main function of HVAC is heating, which is similar to solar radiation. The results of comparison between the PFM and other two methods on a typical day in winter are shown in Fig. 8.

It can be concluded that the temperature and power values obtained based on the PFM method are more in line with the actual values, and thus the accuracy rate is higher than that of the TPM and DDM methods.

Figure 9 presents the box plot of accuracy comparison of the PFM and other two methods on a typical day in winter. In the calculation example, 100 sets of sample data are used to verify the accuracy results. The accuracy of the PFM method is approximately 92.6%, which is clearly higher than that of the other two methods, and it is relatively stable. The TPM method has the lowest accuracy and the worst scalability. The DDM method has the medium calculation accuracy and scalability with a small amount (100 to 300 days) of sample data.

Finally, we present the accuracy comparison of the PFM and other two methods in the training process, as shown in Fig. 10.

The PFM method has a higher training accuracy and a higher starting point for training accuracy. This is because the PFM method is based on an accurate physical model and thus has a high training accuracy from the beginning. The final training accuracy of TPM method is low at only 70% to 80%. However, due to the inherent characteristics of the physical model, the TPM methods could also have a certain initial training accuracy. Because the DDM method is completely based on data, the initial training accuracy is 0, but the DDM method shows a good learning effect and thus the final accuracy is basically the same as that of the PFM method. However, because of its dependence on sample data, the DDM method could not achieve high training accuracy when the amount of training data is relatively small. As shown by the dotted blue line, when the training accuracy of PFM method reaches 90%, the accuracies of the TPM and DDM methods are less than 70% and 80%, respectively.

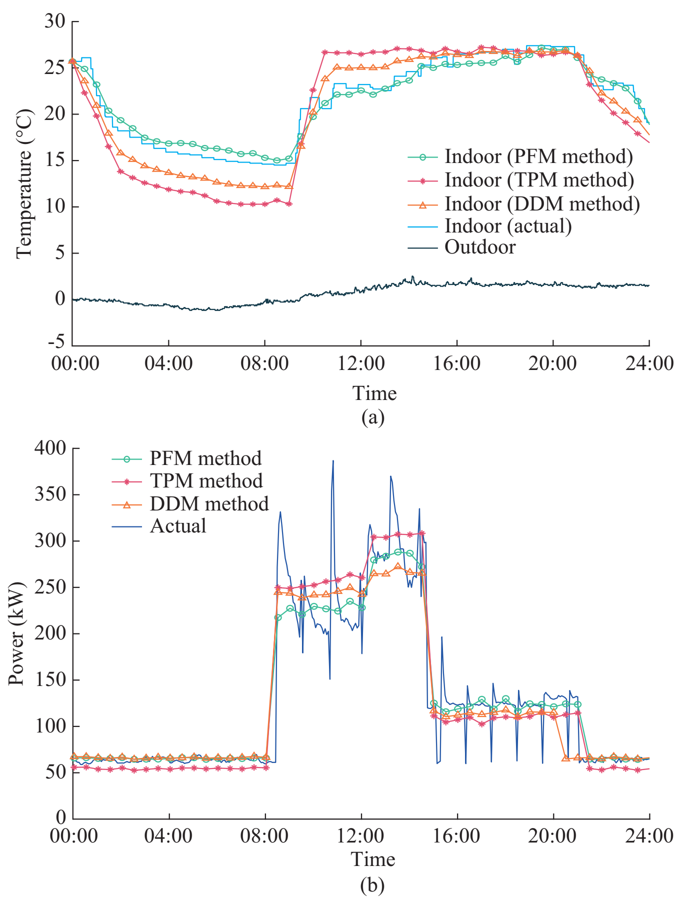


Fig. 8. Results of comparison between PFM and other two methods on a typical day in winter. (a) Comparison of regional temperature. (b) Comparison of regional energy consumption.

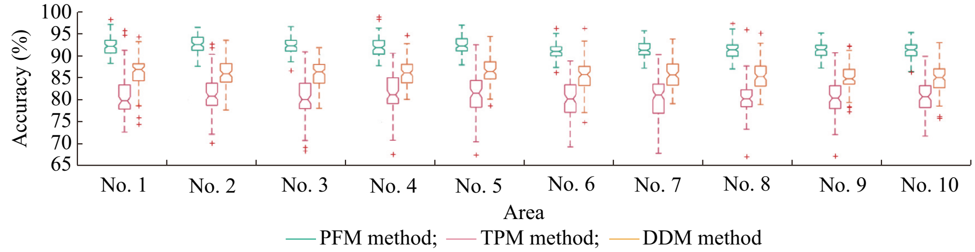


Fig. 9. Box plot of accuracy comparison between PFM and other two methods on a typical day in winter.

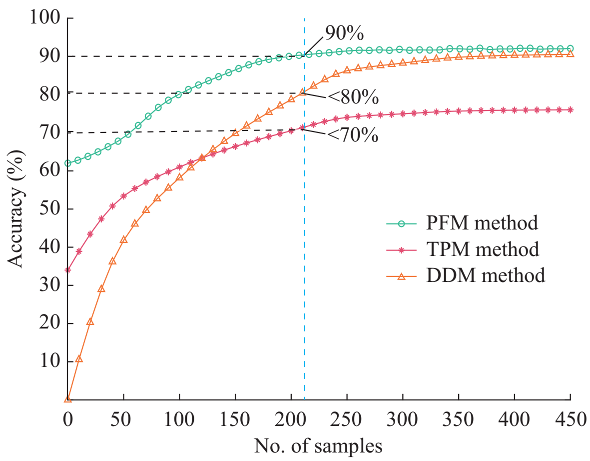


Fig. 10. Accuracy comparison between PFM and other two methods in training process.

## V. Conclusion

An office building in a high-tech park is used as an example to construct a precision physical model of the building. A PFM method is proposed to correct the accuracy of the model through the parameter and model correction. The relevant conclusions are as follows.

1) Existing studies on the architectures for energy consumption analysis have given little attention to precise physical models. This paper constructs a refined physical model for energy consumption analysis of buildings. This model describes in detail the structural matrix and the thermal and thermal-electric conversion processes of a building through physical modeling. In particular, the interaction between the interior areas of the building and behavioral information is considered in the form of a structural matrix.

2) This paper proposes a method for analyzing energy consumption of building through the PFM method. The accuracy of the energy consumption analysis could be improved by modifying the parameters and model. The interactive mechanism of energy conversion in buildings could be retained through a physical model, and the accuracy of the energy consumption analysis could be improved by the DDM method.

3) The energy consumption analysis based on the PFM method proposed in this study can obtain higher accuracy (over 90%) when the sample data volume is relatively small. The problem that the DDM method for building energy consumption analysis of buildings requires a large amount of sample data is solved.

## References

[1] J. Pan, R. Jain, S. Paul *et al.*, "An internet of things framework for smart energy in buildings: designs, prototype, and experiments," *IEEE Internet of Things Journal*, vol. 2, no. 6, pp. 527-537, Mar. 2016.

[2] H. Haider, O. See, and W. Elmenreich, "A review of residential demand response of smart grid," *Renewable and Sustainable Energy Reviews*, vol. 59, pp. 166-178, Jun. 2016.

[3] D. Kolokotsa, "The role of smart grids in the building sector," *Energy and Buildings*, vol. 116, pp. 703-708, Mar. 2016.

[4] D. Zhang, S. Evangelisti, P. Lettieri *et al.*, "Economic and environmental scheduling of smart homes with microgrid: DER operation and electrical tasks," *Energy Conversion and Management*, vol. 110, pp. 113-124, Feb. 2016.

[5] A. F. Taha, N. Gatsis, B. Dong *et al.*, "Buildings-to-grid integration framework," *IEEE Transactions on Smart Grid*, vol. 10, no. 2, pp. 1237-1249, Oct. 2017.

[6] X. Jin, J. Wu, Y. Mu *et al.*, "Hierarchical microgrid energy management in an office building," *Applied Energy*, vol. 208, pp. 480-494, Dec. 2017.

[7] M. Razmara, M. Maasoumy, M. Shahbakhti *et al.*, "Optimal exergy control of building HVAC system," *Applied Energy*, vol. 156, pp. 555-565, Oct. 2015.

[8] M. Razmara, G. R. Bharati, D. Hanover *et al.*, "Building-to-grid predictive power flow control for demand response and demand flexibility programs," *Applied Energy*, vol. 203, pp. 128-141, Oct. 2017.

[9] W. Labeeuw, J. Stragier, and G. Deconinck, "Potential of active demand reduction with residential wet appliances: a case study for Belgium," *IEEE Transactions on Smart Grid*, vol. 6, no. 1, pp. 315-323, Jan. 2015.

[10] M. Brenna, M. C. Falvo, F. Foiadelli *et al.*, "From virtual power plant (VPP) to sustainable energy microsystem (SEM): an opportunity for buildings energy management," in *Proceedings of 2015 IEEE Industry Applications Society Annual Meeting*, Addison, USA, Dec. 2015, pp. 1-8.

[11] X. Jin, Y. Mu, H. Jia *et al.*, "Dynamic economic dispatch of a hybrid energy microgrid considering building based virtual energy storage system," *Applied Energy*, vol. 194, pp. 386-398, May 2017.

[12] T. Jiang, Z. Li, X. Jin *et al.*, "Flexible operation of active distribution network using integrated smart buildings with heating, ventilation and air-conditioning systems," *Applied Energy*, vol. 226, pp. 181-196, Sept. 2018.

[13] X. Chen, J. Wang, J. Xie *et al.*, "Demand response potential evaluation for residential air conditioning loads," *IET Generation, Transmission & Distribution*, vol. 12, no. 19, pp. 4260-4268, Sept. 2018.

[14] C. Zhang, S. R. Kuppannagari, R. Kannan *et al.*, "Building HVAC scheduling using reinforcement learning via neural network based model approximation," in *Proceedings of the 6th ACM International Conference on Systems for Energy-efficient Buildings, Cities, and Transportation*, Zhuhai, China, Dec. 2019, pp. 287-296.

[15] X. Jin, J. Wu, Y. Mu *et al.*, "Hierarchical microgrid energy management in an office building," *Applied Energy*, vol. 208, pp. 480-494, Dec. 2017.

[16] M. Razmara, G. R. Bharati, M. Shahbakhti *et al.*, "Bilevel optimization framework for smart building-to-grid systems," *IEEE Transactions on Smart Grid*, vol. 9, no. 2, pp. 582-593, Apr. 2016.

[17] Y. Ye, D. Qiu, X. Wu *et al.*, "Model-free real-time autonomous control for a residential multi-energy system using deep reinforcement learning," *IEEE Transactions on Smart Grid*, vol. 11, no. 4, pp. 3068-3082, Feb. 2020.

[18] H. Nezamabadi and V. Vahidinasab, "Market bidding strategy of the microgrids considering demand response and energy storage potential flexibilities," *IET Generation, Transmission & Distribution*, vol. 13,

no. 8, pp. 1346-1357, Jan. 2019.

[19] F. U. M. Ullah, A. Ullah, I. U. Haq *et al.*, "Short-term prediction of residential power energy consumption via CNN and multi-layer bi-directional LSTM networks," *IEEE Access*, vol. 8, pp. 123369-123380, Dec. 2019.

[20] J. Song, G. Xue, Y. Ma *et al.*, "An indoor temperature prediction framework based on hierarchical attention gated recurrent unit model for energy efficient buildings," *IEEE Access*, vol. 7, pp. 157268-157283, Oct. 2019.

[21] L. Yu, Y. Sun, Z. Xu *et al.*, "Multi-agent deep reinforcement learning for HVAC control in commercial buildings," *IEEE Transactions on Smart Grid*, vol. 12, no. 1, pp. 407-419, Jun. 2020.

[22] L. Yu, W. Xie, D. Xie *et al.*, "Deep reinforcement learning for smart home energy management," *IEEE Internet of Things Journal*, vol. 7, no. 4, pp. 2751-2762, Sept. 2019.

[23] S. Biao, B. L. Peter, Q. Jia *et al.*, "Building energy management: integrated control of active and passive heating, cooling, lighting, shading, and ventilation systems," *IEEE Transactions on Automation Science and Engineering*, vol. 10, no. 3, pp. 588-602, Jul. 2013.

[24] Z. Xiangyu, P. Manisa, C. Tao *et al.*, "An IoT-based thermal model learning framework for smart buildings," *IEEE Internet of Things Journal*, vol. 7, no. 1, pp. 518-527, Jan. 2020.

[25] S. Cui and J. Xiao, "Game-based peer-to-peer energy sharing management for a community of energy buildings," *International Journal of Electrical Power & Energy Systems*, vol. 123, pp. 1-10, Dec. 2020.

**Xiao Han** received the B.S. degree in electrical engineering and automation and the M.S. degree in electrical engineering from Nanjing Normal University, Nanjing, China, in 2012 and 2015, respectively. He is now pursuing for the Ph.D. degree in electrical engineering in Nanjing University of Aeronautics and Astronautics, Nanjing, China. His current research interest includes smart energy system of buildings.

**Chaohai Zhang** received the B.A., M.S., and Ph.D. degrees from Harbin Institute of Technology (HIT), Harbin, China, in 1985, Navy Aeronautical Engineering Academy, in 1997, and Hong Kong Polytechnic University, Hong Kong, China, respectively. Currently, he is a Professor at Nanjing University of Aeronautics and Astronautics, Nanjing, China. His research interests include electrical discharges, plasma, electrical environment and condition monitoring, and diagnosis of electric power equipment.

**Yi Tang** received the Ph.D. degree from the Harbin Institute of Technology, Harbin, China, in 2006. He is currently a Professor with Southeast University, Nanjing, China. His research interests include smart grid, power system security, power system stability analysis, renewable energy system, and cyber-physical system.

**Yujian Ye** received the B.Eng. degree in Electrical and Electronic Engineering from Northumbria University, Newcastle Upon Tyne, U.K., in 2011, the M.Sc. and the Ph.D. degrees from Imperial College London, London, U.K., in 2013 and 2017, respectively. He is currently an Associate Professor at Southeast University, Nanjing, China, and a Visiting Researcher at Imperial College London. His current research interests include development and application of artificial intelligence techniques in modeling, analysis, and control of power and energy system, and optimization of economics of power system operation and planning.