Energy & Buildings 268 (2022) 112185

Contents lists available at ScienceDirect

Energy & Buildings

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

Generation of sub-item load profiles for public buildings based on the conditional generative adversarial network and moving average method

Yuhang Zhang [a,b,1], Xiangtian Deng [a,b,1], Yi Zhang [a,b,c], Yi Zhang [a,b,*]

[a] Tsinghua-Berkeley Shenzhen Institute, Tsinghua University, Shenzhen 518055, PR China

[b] Institute of Future Human Habitats, Tsinghua Shenzhen International Graduate School, Tsinghua University, Shenzhen 518055, PR China

[c] Department of Automation, Tsinghua National Laboratory for Information Science and Technology (TNList), Tsinghua University, Beijing 100084, PR China

ARTICLE INFO

Article history:
Received 5 December 2021
Revised 8 May 2022
Accepted 14 May 2022
Available online 20 May 2022

Keywords:
Sub-item load profile generation
Supporting specific conditions
Conditional Generative Adversarial Network (CGAN)
Moving average method

ABSTRACT

Increasing research on data-driven methods to optimize energy systems and power grid operation requires a large amount of data with regard to building energy consumption profiles; owing to the difficulty in availing load data, time-consuming collection and the privacy issues in the collection process have become limitations and previous research on total load generation cannot meet research requirements of refined energy control and optimization. In this study, we propose a novel approach based on the conditional generative adversarial network (CGAN) and moving average method to generate sub-item load profiles of building energy consumption to solve the aforementioned problems. Sub-item load profiles include light and socket load, HVAC (Heating Ventilation and Air Conditioning) load, impetus load and special load. The CGAN algorithm is employed to generate sub-item load profiles considering specific conditions (multiple labels) e.g. time, weather, and load shape labels. In addition, the moving average method was used to reduce noise in the generated profiles. The case study was conducted based on real-world sub-item load data collected from office buildings, commercial buildings, and hospitals in Shenzhen, China. We validated the generation performance of the sub-item load profile of CGAN-MA by comparing it with the traditional load profile generation method GAN and variational autoencoder based on three aspects: similarity, variability and diversity. Compared with the traditional model, the proposed model improves the similarity and variability by about 5.7% to 64.8%, 76.7% to 135.5% respectively, and can satisfy the requirements of diversity with the diversity indicator of four sub item generated load is 1.36, 1.93, 1.81 and 2.08 respectively. Furthermore, we compared the generated load and real load possibility distributions under the selected conditions. The results show that the load generated by CGAN-MA is higher on working days, rainy days and hot days than on non-working days, sunny days and cool days, which correspond to the real circumstances, and sub-item B (HVAC) is the most sensitive one to different conditions. The proposed model can be applied to effectively generate sub-load profiles under the required conditions and further help in studies related to the development of data-driven methods for energy consumption prediction, demand-side management and the optimization of power grid operation.

1. Introduction

Building energy consumption majorly contributes to overall energy consumption. It would increase by 70% within 2050 [1]. To achieve energy-saving and intelligent management, several methods and techniques e.g. energy consumption forecasting [2], demand-side management (DSM), demand response (DR) [3], dynamic pricing [4], and intelligent energy systems [5], have been studied in recent years. For example, in incentive-based demand response program, the direct load control is conducted by power companies through giving incentives to customers based on their electric load profiles. Therefore, the customers adjust their energy consumption periods, balancing the supply-demand and maintaining the grid stability [6]. The performance of the energy management system can be measured by analyzing the building electrical load profile to develop the optimal operational control strategies [7]. Meanwhile, with the development of artificial intelligence (AI) and big data techniques, data-driven methods have gained significant traction in the aforementioned research areas

* Corresponding author at: Tsinghua Shenzhen International Graduate School, Tsinghua University, Shenzhen 518055, PR China.

E-mail address: zy1214@sz.tsinghua.edu.cn (Y. Zhang).

[1] This is the first author footnote.

https://doi.org/10.1016/j.enbuild.2022.112185
0378-7788/

Nomenclature

CGAN	Conditional Generative Adversarial Network
MA	Moving average
GAN	Generative adversarial network
MLP	Multilayer perceptron
VAE	Variational autoencoder
HVAC	Heating ventilation and air conditioning
DSM	Demand side management
DR	Demond response
KL divergence	Kullback-Leibler divergence
MAE	Mean absolute error
RMSE	Root mean squared error

[8,9]. For example, the convolutional neural network (CNN) is one of the most basic neural networks which can extract features from deep neural networks. It has been widely applied in machine learning tasks in computer vision [10]. Academia and IT communities have paid considerable attention to big data which is a potential research area recently. Big data technologies can assist in cost reduction, better decision making, product recommendation, fraud detection, etc. The data volume is very important to realize these technologies. For example, intelligent load forecasting can assist in energy optimization, providing operational methods to both energy consumers and producers. Many techniques are used to realize intelligent load forecastings, such as statistical learning, genetic algorithm, neural network, and hybrid methods. However, implementing intelligent load forecasting requires a large number of data sets to train models [11]. Therefore, in some scenarios, in order to better apply big data technology, data sets in corresponding fields should be generated, which can be regarded as data enhancement [12] (See Table 1).

However, data-driven methods face two major limitations: data collection [13] and privacy security [14]. For data collection, smart meters are deployed to collect fine-grained electricity consumption data [15] and send information to a remote server for storage and analysis [16]. However, installation and maintenance costs are relatively too high for users. Moreover, the data collected can be complex, and they require a considerable effort for preprocessing, which implies that some amount of data may be missing or contain outliers caused by the unknown noise. For privacy and security, people are concerned about potential private information leakage if their meter data are shared and manipulated for illegitimate purposes. Therefore, the collected data are limited for utility companies or research institutes [16]. Considering the aforementioned problems, researchers have been investigating methods to generate new data for training. In the past, researchers usually conduct physical modeling of buildings to simulate the electric load profiles. It needs to predetermine some complex parameters, such as the occupancy of the building and the electrical appliance schedules [17]. However, this requires a certain professional background and a detailed investigation of the power consumption behavior of the building [18]. Nowadays, generating load profiles by learning latent stochastic data distribution based on the real data collected is more widely applied. Intuitively, probabilistic generative models are appropriate for this scenario. In [19], the median filter is used to extract an accurate mathematical model of household electric power load, the filtered data is used to estimate model parameters. In [20], Kernel Independent Component (KIC) analysis algorithm is used to generate load profile through power flow. In [21] probabilistic modelling is used to estimate the system-wide PEV charging load in domestic grids. In [22], Variational Auto-Encoder (VAE) was adopted to learn a generative model for electric vehicle (EV) profiles. One of the studies [23] utilized a flow-based model, known as the Nonlinear Independent Component Estimation (NICE) model, to capture the accurate probabilistic distribution of nonlinear time characteristics of the real samples. Meanwhile, in recent years, the Generative Adversarial Network (GAN) has achieved great success in generating image data to train deep neural network models [24]. In [25], GAN is used to generate multispectral satellite images. The result shows the spectral signatures of generated images is similar to that kind of terrain in the images. In addition, It is incorporated into load profile generation. Another study [16] proposed a GAN-based method to generate load profiles for individual buildings. The authors claimed that their proposed model can learn dynamic behaviors of building loads and anonymize user-sensitive information. Moreover, an Auxiliary Classifier GAN (ACGAN) based model to generate differential load profiles

Table 1
Summary of the reviewed models for generation.

Ref No.	Ref	Year	Algorithm	Remarks	Whether consider sub-items and external conditions
[19]	Performance evaluation of median filter in modeling household electric load profile	2018	Median filter	Generate household electric load profile	Not consider sub-items and external conditions
[20]	Electric power load profile estimation applying kernel independent component analysis	2008	KIC	Estimate electric load profile	Consider the consumption patterns but not consider the external conditions
[21]	Probabilistic estimation of plug-in electric vehicles charging load profile	2015	Monte Carlo simulation	Estimate electric vehicles charging load profile	Not consider sub-items and external conditions
[22]	Data-driven ev load profiles generation using a variational auto-encoder	2019	VAE	Generate electric vehicle profiles	Not consider sub-items and external conditions
[23]	Modeling daily load profiles of distribution network for scenario generation using flow-based generative network	2020	NICE	Generate daily load profiles of distribution network	Not consider sub-items and external conditions
[25]	.Gan generation of synthetic multispectral satellite images	2020	GAN	Generate multispectral satellite images	Not consider sub-items and external conditions
[16]	Generating realistic building electrical load profiles through the generative adversarial network	2020	GAN	Generate electric load profiles	Not consider sub-items and external conditions
[26]	Gan-based model for residential load generation considering typical consumption patterns	2019	ACGAN	Generate differential load profiles	Consider the consumption patterns but not consider the external conditions
[27]	Conditional tabular GAN-based two-stage data generation scheme for short-term load forecasting	2020	CTGAN	Predict short term load	Not consider sub-items but consider external conditions

was presented [26]. This model performs an extra classification task based on the GAN model to learn data distribution to classify different typical residential load patterns. [27] proposed Conditional Tabular GAN (CTGAN) based generation method for short-term load forecasting. Most researchers have studied the overall load profile and few have considered the sub-item load profile. However, refined management and control must manage a more fine-grained load to improve performance. Sub-items with different energy levels and patterns will heterogeneously impact refined management and control. The time sequence of the building electric load profile is complex and affected by both internal and external factors [28]. Different sub-item electric load profiles have different temporal features and present different patterns. The flexibility of different sub-item electric load profiles under internal and external conditions is also different. For example, the sequence of the socket load profile is stationary which is mainly affected by external factors while the sequence of HVAC (Heating Ventilation and Air Conditioning) load profile is non-stationary which is more uncertain and affected by both external factors and internal factors [29]. They should be simulated separately [30]. Some recent energy-related application research including energy price prediction, energy management and demand response also try to recognize the different patterns of electric load profiles. For example, in [31], due to the inherent non-stationarity characteristics of energy prices, the independent component analysis is used to recognize the inherent features of energy price, further improving the accuracy of prediction. In [32], a driving pattern recognition method is proposed to optimize the energy management for plug-in hybrid electric buses. In [33], it is suggested that the demand response strategies development for specific loads like HVAC, EVs, computer severs, etc gradually become a new research direction. In addition, although aforementioned studies have proved that generative models including probabilistic models and GAN are promising to generate load profiles, which appear real and maintain dynamic behaviors of the load profiles. Few studies have considered the connections between load profiles and some explicit variables. These variables represent the building environment (e.g. air temperature), temporal variables (e.g. weekdays) and load shape characteristics. There are some benefits if we can generate new load profiles as required. For example, a sufficient number of load profiles under different required conditions needs to be generated to evaluate a new energy management strategy and the simulation of sub-item load profiles under specific conditions can better serve the refined management, control and optimization. To fill this research gap, we propose a conditional GAN (CGAN) based method. The key contributions of this paper are as follows:

- We consider that different sub-items present different patterns and are affected by external factors to varying degrees. CGAN-MA model is proposed to generate the sub-item load profile under specific conditions for buildings which can assist in refined energy management and control.
- Multiple conditions are introduced including load profile characteristics, time conditions and weather conditions, providing supervision information to guide the model to generate electric load profiles under corresponding conditions. In addition, the noise in the generating process is reduced by integrating CGAN with the moving average method.
- The case study is conducted over four sub-item load profiles respectively in office building, commercial building and hospital in Shenzhen, China. The proposed model can not only simulate the sub-item electric load profile under different day types, weather conditions, and temperatures but also have a better generation performance than the general methods in the aspect of similarity, diversity and variability.

The rest of the study is organised as follows: Section 2 elaborates our proposed CGAN-MA method in detail. Section 3 presents the dataset, experimental settings. Section 4 presents the results of the case study. Section 5 discusses the related content. Section 6 concludes the study.

2. Methodology

2.1. Framework

Fig. 1 shows the framework of the proposed method that generates sub-items and overall building load profile under specific conditions. This includes three major steps:

1. Data Preprocessing. The total load profile is decomposed according to the coding regularization of the sub-item, and then each sub-item load profile is split into daily granularity and normalized.
2. Condition Construction. We exact load shape label (peak load, base load, high load start time, high load finish time, high load duration), time label (month, day, day type), and weather label (highest temperature, lowest temperature, weather, wind).
3. Load Profile Generation. The multi-condition and seeds randomly sampled from predetermined distribution are concatenated and input to the conditional generative adversarial neural network with the moving average method(CGAN-MA).

In this study, we combine the conditional GAN (CGAN) based on past information. The time granularity of the load profile is 15-min. We smooth data points produced by the generator using the moving average method, which can reduce noise, allowing the generated load better learn the time series characteristic.

2.2. Multiple label condition generation

2.2.1. Load shape condition

In this study, the load shape is considered as an important label for load profiles, which represents variations in load including the trend and randomness. The load profile characteristic depicts the load shape of a building during a specific period and can reflect different dynamic and stochastic details of the load for different types of buildings [34]. For example, the load level of an office building is higher than that of a commercial building. Moreover, the load level change at different times of the day. For office buildings, the load increases in the morning, arrives at its peak around noon, and decreases in the afternoon [35]. [36] quantified the characteristic of load shape using peak load, base load, rise start time, high load start time, high load finish time, fall finish time, rise time, high load duration and fall time. The rise start time and fall finish time are calculated based on the base load and peak load, which are affected by the extreme event, while the high load start time and finish time are calculated based on the halfway of the threshold percentile of the load, which is more stable.

In this study, we quantify the periodic variations in the load profile using the following characteristics [36]:

- Peak load: 90 percentile of daily load
- Base load: 10 percentile of daily load
- High load start time: the earliest time when the load is more than halfway to the 90 percentile load for the day
- High load finish time: the final time when the load is more than halfway to the 90 percentile load for the day
- High load duration: the time interval of load between high load start time and finish time.

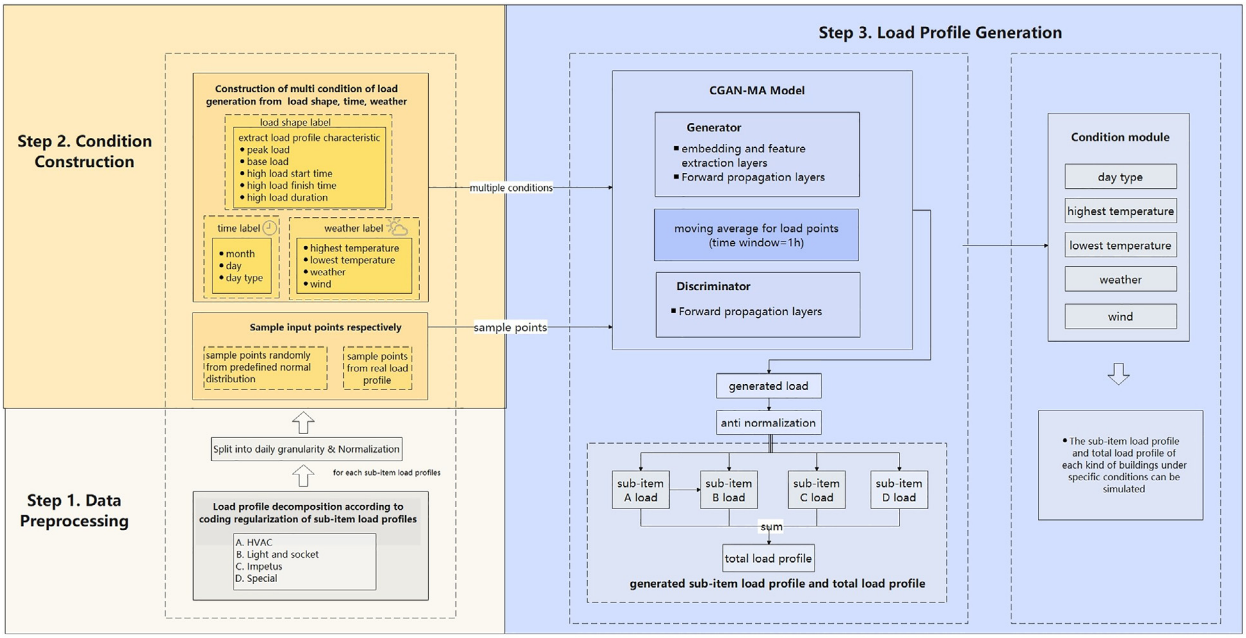


Fig. 1. Methodology Framework.

Based on the distribution of load data, we define the 90 percentile as the peak load and the 10 percentile as the base load to stabilize the key parameters, thereby reducing the interference from outliers.

2.2.2. Time condition

We classified the whole time series load data into daily load data and assumed it is conditional dependent on its date information including month, day and day type. The load level around June and July is usually higher than that in other months. The day type describes whether it is a working day or not. For office buildings, the load level on working days is higher than that on non-working days.

2.2.3. Weather condition

We assumed that the daily load profile is conditionally dependent on meteorology. Weather, wind, and temperatures(highest and lowest) were included in the weather label. The load profiles exhibit different behaviors under different weather conditions [37].

2.3. CGAN-MA Model

In this study, we propose the CGAN-MA model by combining the conditional generative adversarial network and the moving average method. Fig. 2 shows the structure of the CGAN-MA model. The extracted load shape condition, time condition, and weather condition are concatenated with sample points from Gaussian distribution. These concatenated vectors are input into the generator. The output vectors are the original generated load with much noise, and then, they are input into the moving average(MA) module. The module reduces the noise of generated load profile and smooths it to align with the time series characteristics.

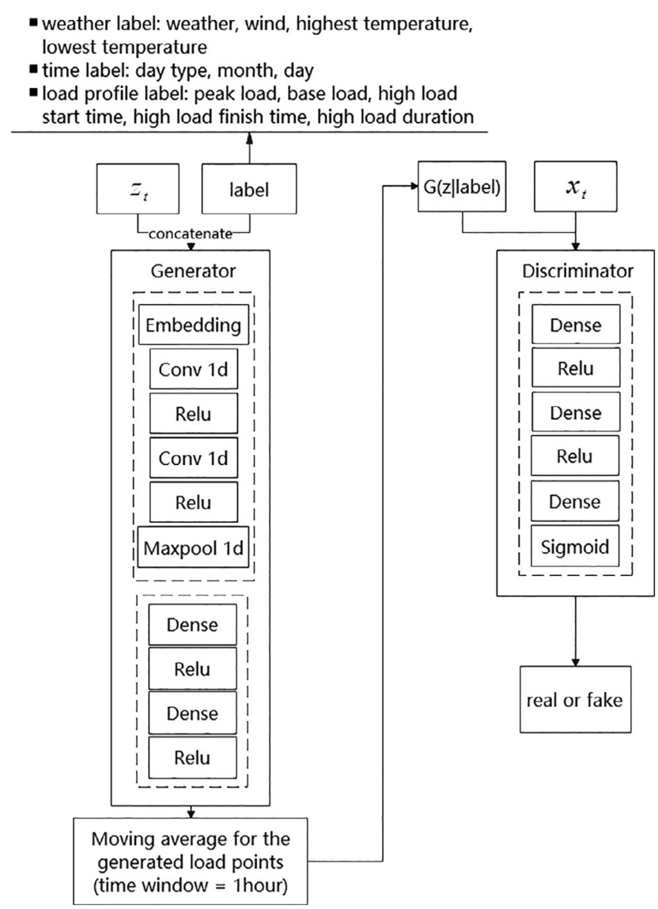


Fig. 2. The structure of CGAN-MA model.

2.3.1. Conditional generative adversarial network

The GAN is one of the most promising models of unsupervised learning for complex distribution in recent years. It has been used to generate images [38] and performs well based on the mutual game learning of two modules in the framework: generator and discriminator. The generator produces new samples whose distribution is similar to real samples but not the same. The objective of the discriminator is to judge whether the sample is fake and try to identify fake samples generated by the generator. The CGAN, which is an extension of the GAN, was proposed by Mirza [39]. The input of the CGAN is conditioned on exogenous variables compared with that of the GAN.

Fig. 2 shows the structure of the generator and the discriminator. In this study, the load shape label, time label, and weather labels were considered as conditions. Points are randomly sampled from Gaussian distribution and conditions are concatenated to a vector and are provided as input to the generator. We use 2-layer 1D convolutional network, maxpool layer and ReLU activation to extract the main part from the original input and feed this vector representation into 2-layer linear network and generate the initial load curve. The convolutional layer can extract features by training a small number of parameters. After feature extraction in the convolution layer, the output feature map will be transferred to the pooling layer for feature selection and information filtering. The pooling layer contains a preset pooling function, which can replace the result of a single point in the characteristic graph with the characteristic graph statistics of its adjacent regions. ReLU is Rectified Linear Unit which is used to increase the nonlinearity of the neural network model [40]. The discriminator includes three linear layers, and it takes not only the real load but also the load generated by the generator and multiple conditions as input. Finally, based on sigmoid activation, the discriminator determines whether the input load curve is real or fake.

In the training process, the generator and discriminator improved through the competition between them. The loss function of the generator is defined as

$$G\_loss = \frac{1}{m}\sum i = 1^{m}\log(1 - D(G(z_i|y))) \tag{1}$$

where $G(z_i|y)$ is the generated load profile under condition y. $D(G(z_i|y))$ is the probability that the discriminator judges the load profile produced by the generator as the real load profile. $1 - D(G(z_i|y)$ is the probability that the generator judges the load profile produced by the generator as the fake load profile. The generator hopes that the probability of the discriminator judging the generated load as a fake load is as low as possible, i.e. the probability of misjudging generated load as a real load is as high as possible to maximize $\log(D(G(z_i)))$ and minimize $1 - \log(D(G(z_i)))$ loss function. For discriminator, the loss function is defined as

$$D\_loss = \frac{1}{m}\sum i = 1^{m}[\log(D(x_i|y) + \log(1 - D(G(z_i|y)))] \tag{2}$$

where $D(x_i|y)$ is the probability that the discriminator judges the real load as real load under condition y. When judging the load profile, we try to ensure that the real load profiles are judged to be real while the generated load profiles are judged to be fake. Therefore, the loss of the discriminator is the sum of these two parts, and the loss should be maximized.

Algorithm 1 shows the training process of this model which includes three steps: (1) sample noise samples from Gaussian distribution and sample real samples from real profiles, and concatenate them with multiple conditions respectively; (2) for each iteration, feed the real samples and generated samples to the discriminator and then train the discriminator; (3) for each iteration, feed the vector representation of noise samples and multiple conditions to the generator and then train the generator. Algorithm 1 shows the training process of this model. It includes three steps: (1) sample noise samples from Gaussian distribution and sample real samples from real profiles, and concatenate them with multiple conditions respectively; (2) for each iteration, feed the real samples and generated samples to the discriminator and then train the discriminator; (3) for each iteration, feed the vector representation of noise samples and multiple conditions to the generator and then train the generator.

Algorithm 1: Training of Conditional Generative Net

for k epochs **do**
sample m noise samples$z_1, \dots, z_m$
output m synthetic load profiles
sample m real load profiles$x_1, \dots, x_m$
For discriminator:
Input m synthetic load profiles, m real load profiles$x_1, \dots, x_m$, and multiple conditions$y$
Train the discriminator to minimize
$-\frac{1}{m}\sum_{i=1}^{m}[\log(D(x_i|y) + \log(1 - D(G(z_i|y)))]$
Output the probability that sample is real sample
For generator:
Input m noise samples$z_1, \dots, z_m$, multiple conditions$y$
Train the generator with moving average to minimize
$\log(1 - D(G(z_i|y)))]$
Output the synthetic load profiles
end for

2.3.2. Moving average

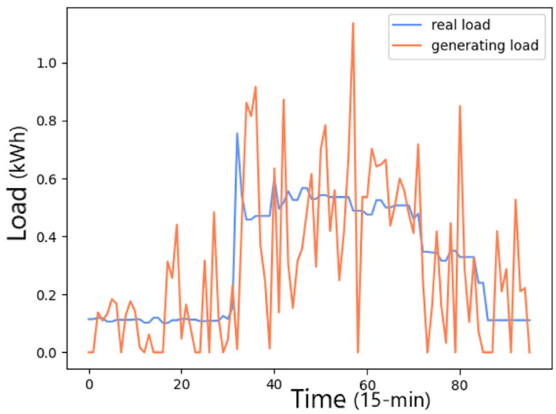
The moving average method is a smoothing prediction technology. It calculates the sequence time average, which is based on a certain number of items according to the time series data, to reflect the trend [41] . Therefore, when the value of the time series significantly fluctuates owing to the influence of periodic variation and random fluctuation, and it is not easy to show the development trend of events, the moving average method can eliminate the influence of these factors and exhibit the trend of the time series. In the CGAN training process, the input of the generator includes a set of noise samples, and the entire process evolves through the game process between the generator and the discriminator. Therefore, even if the overall distribution was gradually optimized, there were several unstable factors; e.g. the impact of the initial input noise still exists. In the whole learning process, the fluctuation range is relatively intense owing to game learning, The difference between points in each sample vector is large, but in practice, as the load data are time-series data, the fluctuation of load data points within 1h is not very volatile. To solve this problem, we considered the time dependence of the load data. In this study, the time granularity of the load profile data was 15-min; therefore, in a day, the load had 96 data points. We perform a moving average on the raw data produced by the generator. The time window of the moving average process was 1-h, i.e. four data points. Fig. 3 compares the raw generated load in CGAN generating process and the generated load after the implementation of the moving average which is in CGAN-MA generating process. It can be found that the noise can be reduced by applying the moving average method and the fluctuation of the generated load is similar to the fluctuation of the real load.

2.4. Generation performance measurement

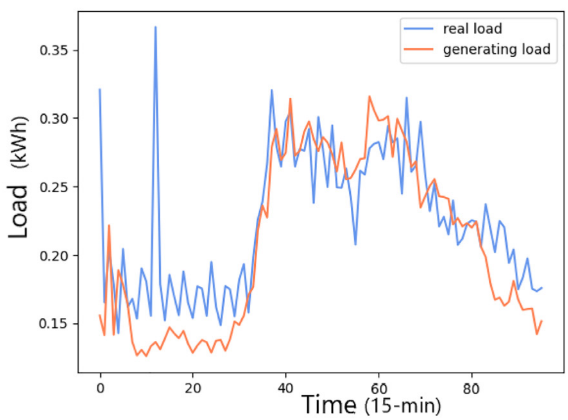
In this study, we validate the generation performance of CGAN-MA based on three aspects: similarity, variability and diversity.

2.4.1. Similarity

Differences in the mean values of key parameters and KL divergence are employed to analyse the similarity. The mean value of key parameters is a statistical indicator of the data distribution. It is defined as follows:



(a) The generated load and real load in CGAN generating process



(b) The generated load and real load in CGAN-MA generating process

Fig. 3. The comparison of CGAN and CGAN-MA generating process.

$$Mean \quad difference = \frac{\sum_{i=1}^{n} x_i\prime - x_i}{n} \tag{3}$$

n is the number of load profiles. x_i' is the key parameter of generated i th daily electric load profile. x_i is the key parameter of real i th daily electric load profile. Kullback-Leibler (KL) divergence [42] is employed as another metric to analyze the similarity, which can measure the difference between two probability distributions. The KL divergence is defined as follows:

$$D(P||Q) = \int_x \mathrm{P}(x)\log\left(\frac{P(i)}{Q(i)}\right)\mathrm{d}x \tag{4}$$

P and Q are two distributions. In our study, P is the distribution of key parameters of the real load profile and Q is the distribution of corresponding key parameters of generated load profile, and x represents key parameters.

2.4.2. Variability

In the process of training the generation model, there is a phenomenon in which similar types of samples are generated [25], leading to the possibility that some statistics are similar overall, but there are not many differences between each other. The generation of similar samples caused by model collapse is meaningless. Therefore, we use the variability indicator [26] to measure the differences between the generated samples. Variability is defined as the average of the Euclidean distance between two samples in the generated load profile group.

2.4.3. Diversity

The objective of generating load profiles is to obtain a group of new load profiles that are similar to real load profiles based on the statistics of key parameter distributions and distinct from real load profiles that are not the same as real load profiles [43]. The diversity is measured by the average of Euclidean distance between each sample in the generated load profile group and the nearest sample in the real load profile group [44].

3. Case study

3.1. Dataset description

We conducted a case study of three different types of buildings to show the application value of this method. We validate the machine learning efficiency of the generated load profile based on short-term load forecasting for office buildings, commercial buildings and hospitals. The commercial buildings mainly refer to shopping malls. We simulated the sub-item load and total load for different types of buildings under specific periods and weather conditions.

Load profiles of three types of large public buildings (office building, commercial building and hospital) in Shenzhen were considered in this study. The construction areas of these three buildings are 37496, 20000, and 33000 m^2 respectively.

The total load profiles of the three buildings in 2017 are shown in Fig. 4. It can be observed that the level of load profile of the hospital is higher than that of the office building and the level of load profile of the commercial building is the lowest. In the overall trend, the load of the buildings increases first and then decreases.

The level and trend of different sub-item loads are different [45], e.g. the load level of lighting and HVAC is greater than that of other purposes. In addition, for the daily load profile, the load of lighting and HVAC always increases and then decreases while other sub-item load does not increase or decrease significantly in the short term. Therefore, we use the sub-item load profile by decomposing the overall load data for generation. It can improve the performance of load profile generation by considering that different trends and different levels of power load significantly impact the mixed overall load. Furthermore, the main sub-item affecting the overall load trend or level is found. The total load profiles were divided into four sub-item load profiles and coded according to their use. The coding regularisation is listed in Table 2.

Lighting and socket electricity mainly refers to the energy consumption of lighting and sockets, corridors, emergency lighting and outdoor landscape lighting. HVAC electricity refers to the energy consumption of hot and cold stations and air conditioning. Impetus electricity refers to the consumption of different impetus devices, such as actuators, tap water pressurization, and sewage discharge, etc. Special electricity refers to the energy consumption of information centers, laundry rooms, kitchens, dining rooms, swimming pools, gyms and other special areas. As shown in Fig. 5, different sub-item load profiles exhibit various levels and patterns.

3.2. Data preprocessing

We collected weather data, which include the highest temperature, lowest temperature, wind speed, and weather during the corresponding period of load data and coded wind and weather

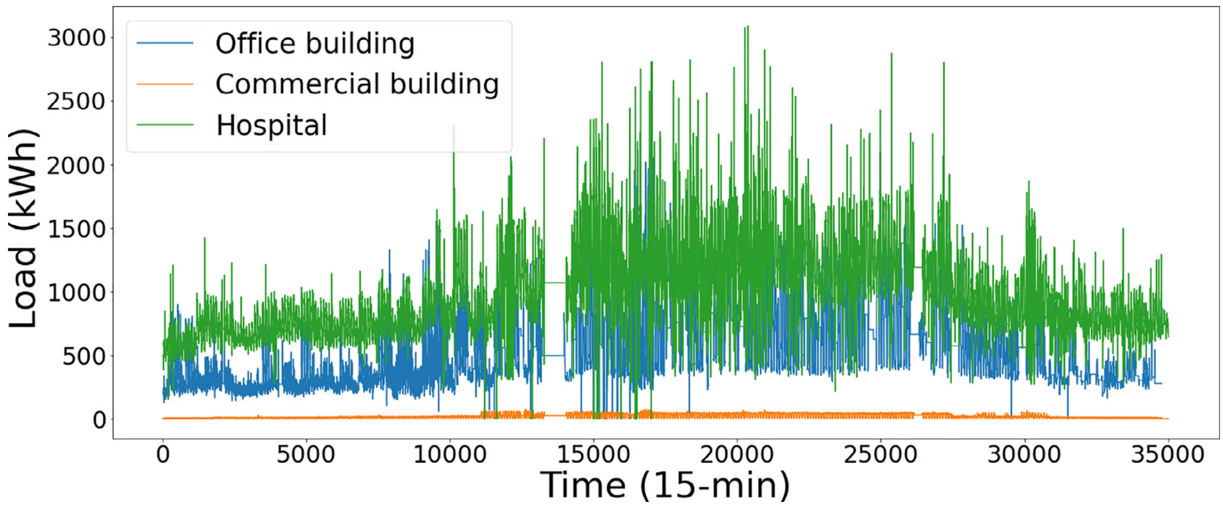


Fig. 4. The load profile for office building, commercial building, hospital.

Table 2
Sub-item load coding regulation.

Type of sub-item load	Code
Light and socket	A
HVAC	B
Impetus	C
Special	D

according to different situations. In addition, we extracted the month, day and day type of daily load profiles. The day type refers to whether the day is a working day or not. One example of the distribution of the sub-item load profile under different times and weather conditions is shown in Fig. 6. The weather conditions were significantly correlated with the load profiles. For example, it was found that the load level on the working day was higher than that on the non-working day. The load level in the middle of the year was higher than that at the beginning and end of the year. The load at noon was higher than that at other times. For the temperature, the higher the temperature, the higher the load level. Therefore, it would be relevant to consider weather information and time information as conditions to feed into the CGAN. In this study, we expect the load profile to be conditioned on the required input weather labels.

To eliminate the difference among sub-item load levels and improve the convergence speed of the model, the load of each sub-item was normalised and used for generated.

3.3. Model verification

3.3.1. Sub-item load profile generation

We implemented the experiment by using PyTorch in Python on a PC with the following configuration: AMD Ryzen 7, 2.9 GHz CPU, 16 GB RMA, and 64-bit system. 'We compared the generation performance of GAN, VAE, and CGAN-MA on all four sub-item load profiles to validate the generation efficiency of CGAN-MA.

Fig. 7 shows the generated sub-item load profile of the office building using the four models. As shown in Fig. 7 the load profile generated by the GAN can present the trend to some extent, but the fluctuation is significantly high. The mean value of the load profile generated using the VAE method is similar to the mean value of the real load profile, but the shape is roughly the same without sufficiently reflecting the randomness of the real load profile enough. The trend of the mean value of the load profile generated using the CGAN-MA method is similar to that of the real load profile, and can reflect the randomness of the load curve. Integrating with the moving average method to consider historical information reduces the influence of noise, so that the generated load profile reflects a trend similar to the original load curve and exhibits dynamic behavior within the random fluctuation range of the real load profile.

We evaluated the similarity, variability and diversity to quantify the performance of generating load profiles using different models.

3.3.2. Similarity

We used the statistics of key parameters(peak load, base load, high-load start time, high-load finish time, high-load duration) of the load shape characteristics [36] to analyse whether the generated sub-item load profile has a distribution similar to that of the real load profile.

First, the difference in the mean value of the key parameters is the metric employed to analyse similarity. It can be found from Fig. 8 that most of the difference of mean value of key parameters of the sub-item load profiles generated by VAE and CGAN-MA is smaller than that of the sub-item load profiles generated by GAN. It can be attributed to the noise input and the training process similar to the game learning without any processing. For the profiles generated by VAE and CGAN-MA, the sum of the mean difference of peak load for four sub-item load profiles generated by CGAN-MA is 22.06 less than that generated by VAE. The sum of the mean difference of base load for four sub-item load profiles generated by CGAN-MA is 20.65 less than that generated by VAE. The sum of the mean difference of high load start time for four sub-item load profiles generated by CGAN-MA is 16.08 less than that generated by VAE. The mean differences of key parameters of the four sub-items are taken as absolute values. The sum of the mean difference of high load end time for four sub-item load profiles generated by CGAN-MA is 0.34 more than that generated by VAE. The sum of the mean difference of high load duration for four sub-item load profiles generated by CGAN-MA is 2.92 less than that generated by VAE. The mean differences of key parameters of the four sub-items are taken as absolute values. In the whole, the mean difference of key parameters of load profile generated by CGAN-MA is less than that generated by GAN and VAE. Additionally, for the profiles generated by VAE and CGAN-MA, the mean difference of the peak load and the base load of sub item B (HVAC) is large, which can be attributed to the continuous

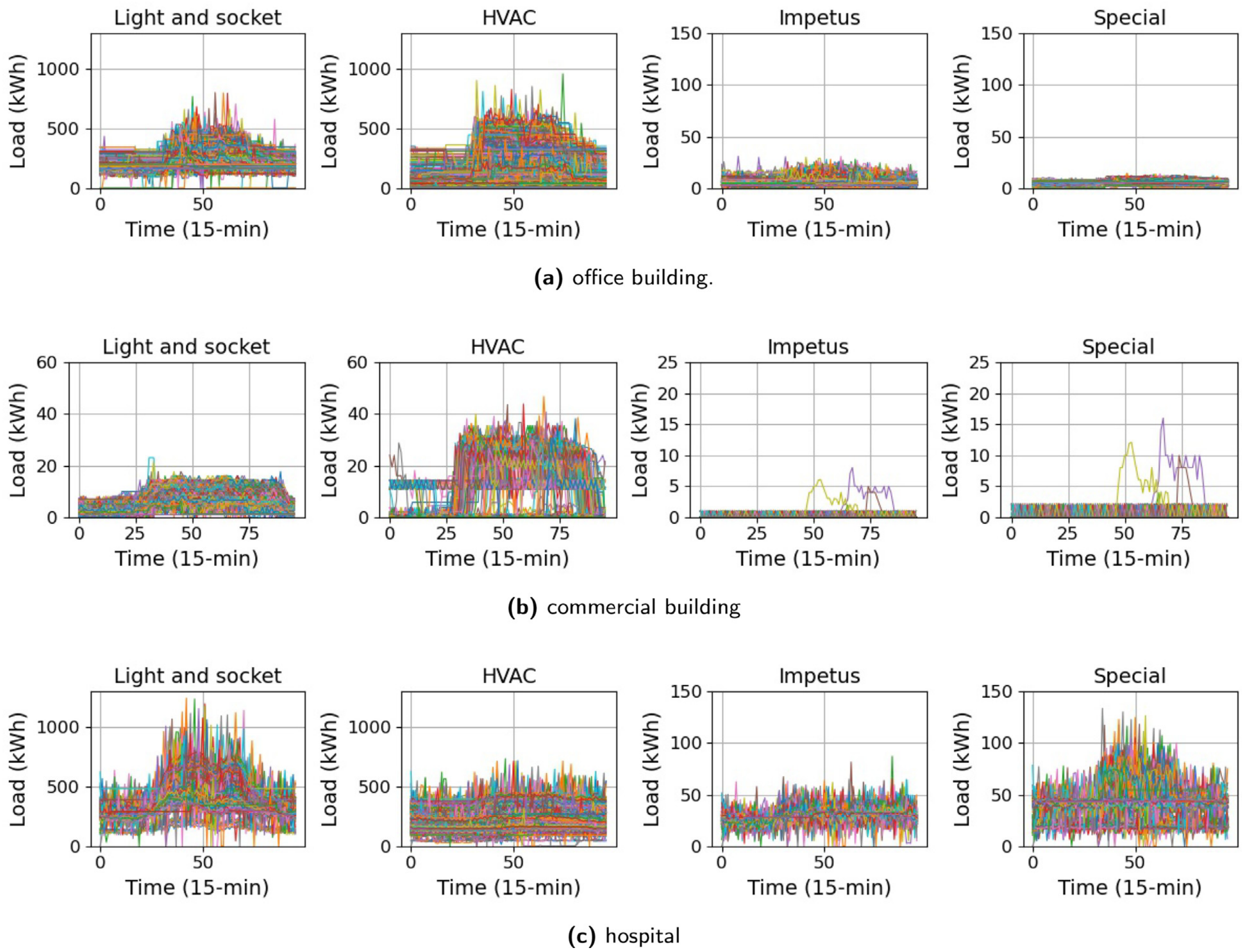


Fig. 5. The daily sub-item load profile for office building, commercial building and hospital.

extreme event. The continuous extreme event includes extreme weather conditions and meter failure. The existence of extreme weather conditions leads to the variation of peak load and base load. The meter failure leads to inaccurate load measurement[46].

In addition, Kullback-Leibler(KL) divergence [42] is employed as another metric to analyze the similarity, which can measure the difference between two probability distributions.

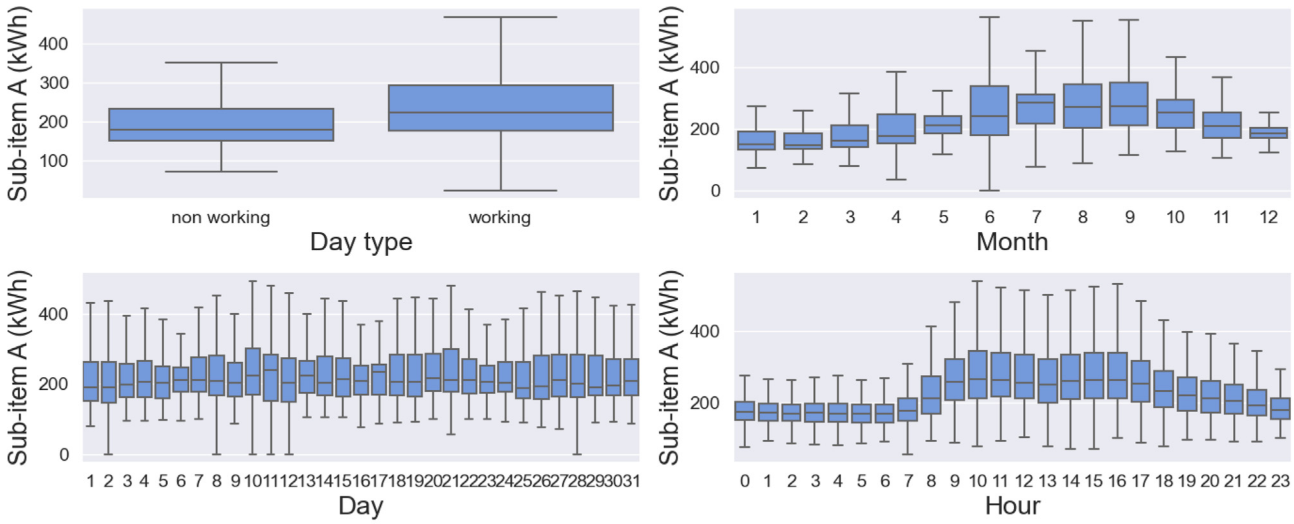
The smaller the KL divergence is, the more identical the two distributions are. Fig. 9 shows the KL divergence of the probability distribution of key parameters for the four sub-items load profiles in comparison. It can be observed that the KL divergence of the majority parameters for each sub-item load profile generated by CGAN-MA is less than 0.3 which is smaller than those generated by other models. Therefore, in terms of similarity, CGAN-MA performs better in terms of the evaluation of the mean difference and KL divergence of the key parameters. In addition, it can be observed that the probability distribution of key parameters of the load profile generated by the GAN is more identical to that of the real load than that of the load profile generated by the VAE.

3.3.3. Variability

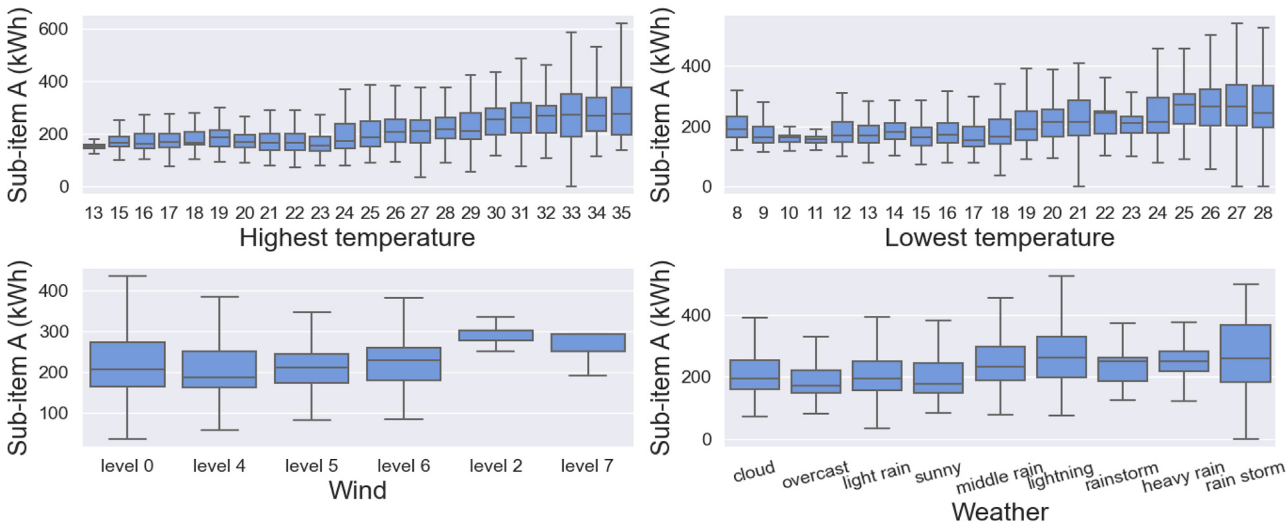
Table 3 shows the variability of real load profiles and load profiles generated by different models. The baseline is the real load profiles. It can be observed that the variability of sub-item load profiles generated by GAN and VAE is small which can be attributed to the mode collapse in the generation process leading to generating similar load curves constantly. CGAN-MA can generate various types of load profiles with the variability indicator for four sub-item being 3.20,9.95,4.28,5.51. The variability indicator of sub-item A load profile generated by CGAN-MA is 1.3 higher than that generated by GAN, 2.45 higher than that generated by VAE. The variability indicator of sub-item B load profile generated by CGAN-MA is 4.28 higher than that generated by GAN, 2.78 higher than that generated by VAE. The variability indicator of sub-item C load profile generated by CGAN-MA is 1.47 higher than that generated by GAN, 2.59 higher than that generated by VAE. The variability indicator of the sub-item D load profile generated by CGAN-MA is 2.89 higher than that generated by GAN, 0.98 higher than that generated by VAE. In addition, the value of variability of four sub-item load profiles generated by CGAN-MA is closer to the baseline. Therefore, CGAN-MA has a better performance than GAN and VAE in the aspect of variability which can not only generate various types of load profiles but also control the randomness and fluctuation in a range similar to that of the real load profile.

3.3.4. Diversity

Table 4 lists the results of the diversity of the load profiles generated using different models. We want to analyze whether the score of diversity is a very low number close to zero which means that the generated load profiles and real load profiles are almost the same. We observe that the load generated by these four methods is distant from real load profiles, which avoids the problem of generating a group of load profiles similar to the real load profiles. VAE scores were the lowest for sub-items B, C, and D, and the score in the evaluation of the variability indicator was relatively small. Therefore, it can be inferred that the VAE may find some type of



(a) The distribution of sub-item load under different time conditions for office building



(b) The distribution of sub-item load under different weather conditions for office building

Fig. 6. The distribution of sub-item load.

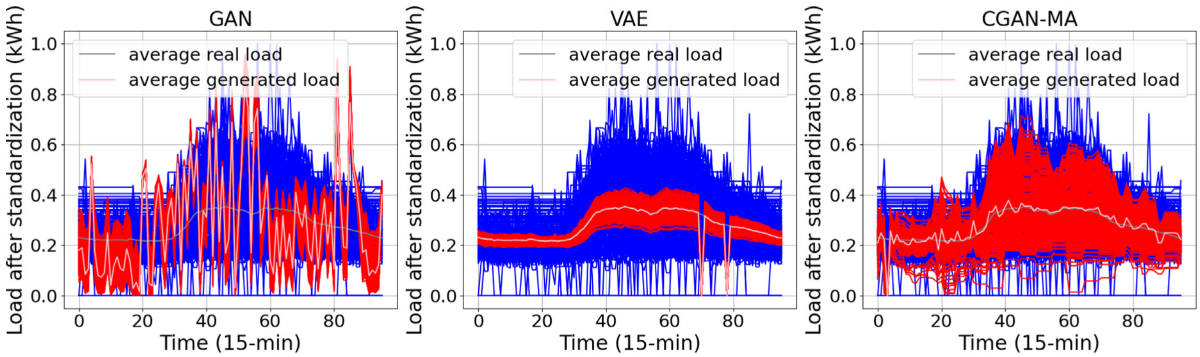
load in the real load profiles and continuously generate this type of load within a certain range around the overall mean value and perform well in the evaluation of the similarity of the key parameter mean value indicator. However, the evaluation of the variability and diversity indicators of the overall distribution is poor.

The principles of GAN and CGAN-MA are roughly the same. These principles of them all belong to the game evolution between the generator and discriminator to generate a batch of curves. With regard to the evaluation of diversity, the CGAN-MA score was relatively small but greater than that of VAE. Neither an extremely small score nor an extremely large score on the diversity indicator means a better generation performance because of the randomness and fluctuation. Therefore, CGAN-MA can satisfy our requirements in terms of diversity while GAN obtains a high score in diversity by sacrificing the similarity of statistics of key parameters with real load profile owing to over randomness and fluctuation.

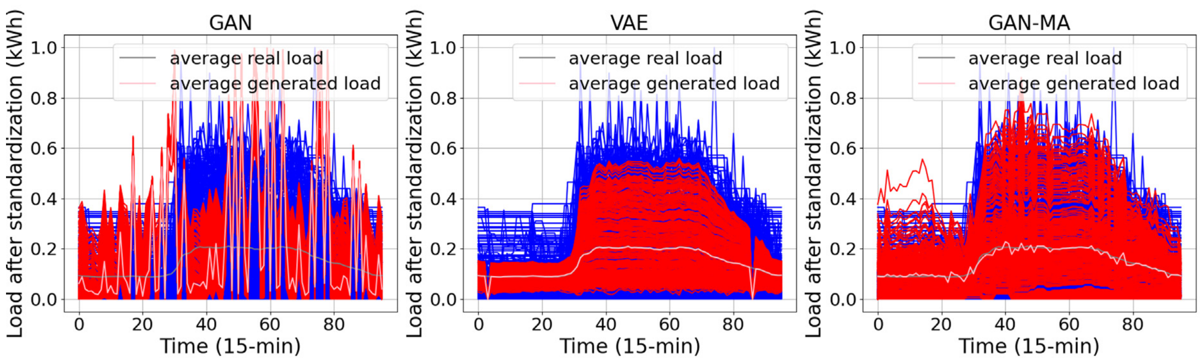
In addition, we plot the mean and standard deviation of the generated load profiles using different models respectively to visualize the diversity of the generated load profiles and analyze whether the generated load profiles lose diversity. As shown in Fig. 10 that the mean and standard deviation of the load profiles generated by GAN and VAE are relatively concentrated, and the range of the mean is similar. However, the standard deviation of the GAN method is much larger than that of the real load profile distribution, and the volatility and randomness are significant. The relationship between the standard deviation and the mean of the VAE method is defined by a curve, while the scatter distribution of the mean and standard deviation of the real load profile is scattered and there is no obvious relationship, and there are more possibilities. The mean-standard deviation of the load profile generated by CGAN-MA is relatively scattered, which is similar to that of the real load profile. The distribution area of the mean-variance point distribution of the load profile generated by CGAN-MA coincides with the area of the mean standard deviation point distribution of the real load profile which indicates that it can simulate a possible sub-item load profile, similar to the real load profile.

3.3.5. Machine learning efficiency

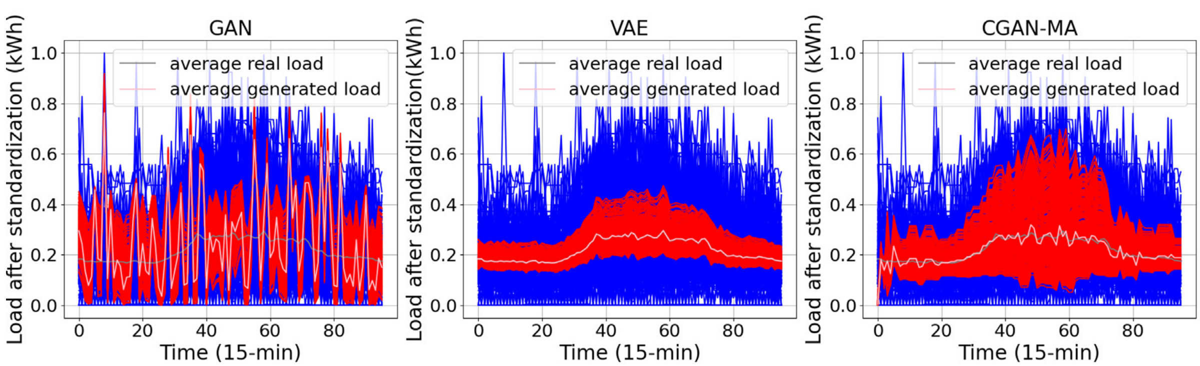
The objective of generating sub-item load profiles is to generate a batch of load profiles that can be used for energy management,



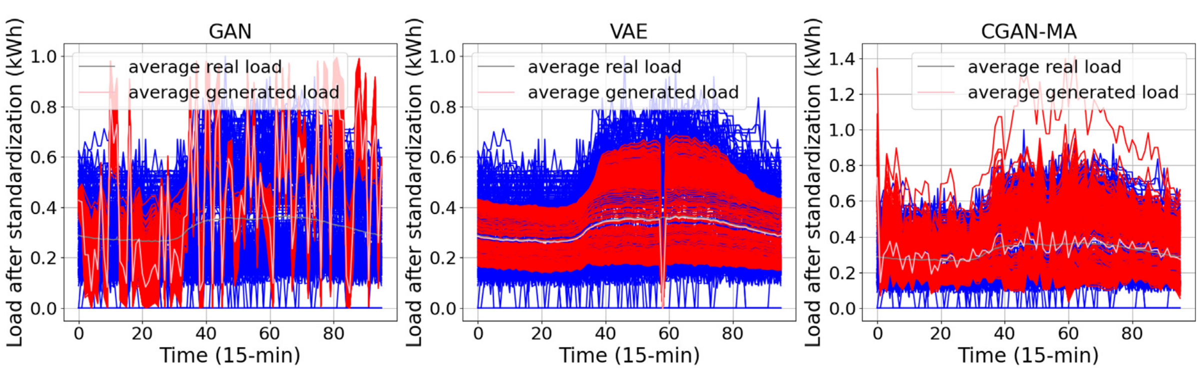
(a) Sub-item A load profile generated by GAN,VAE,CGAN-MA



(b) Sub-item B load profile generated by GAN,VAE,CGAN-MA



(c) Sub-item C load profile generated by GAN,VAE,CGAN-MA



(d) Sub-item D load profile generated by GAN,VAE,CGAN-MA

Fig. 7. The four sub-item load profile generated by GAN, VAE, CGAN-MA for office building.

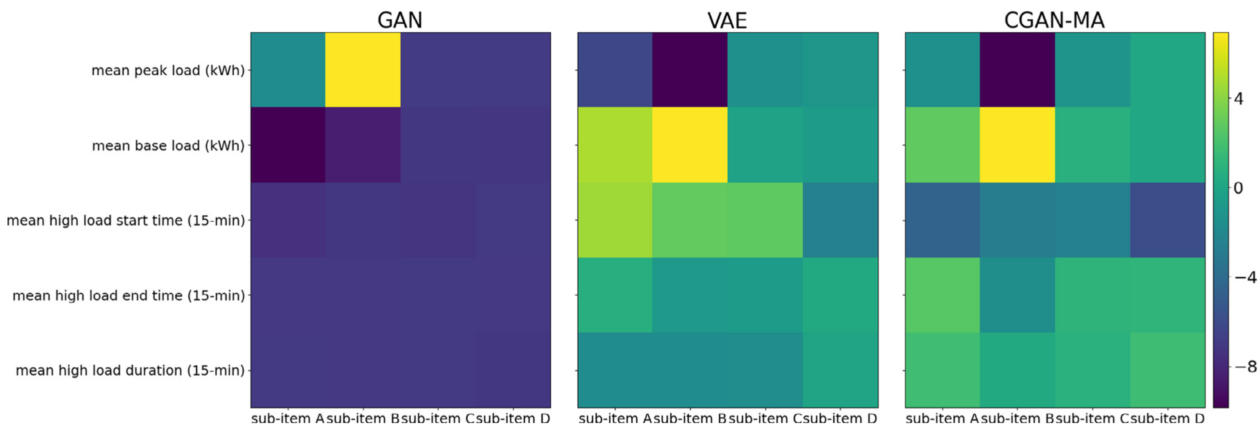


Fig. 8. The mean difference of key parameters of four sub-item load profiles generated by GAN, VAE, CGAN-MA for office building.

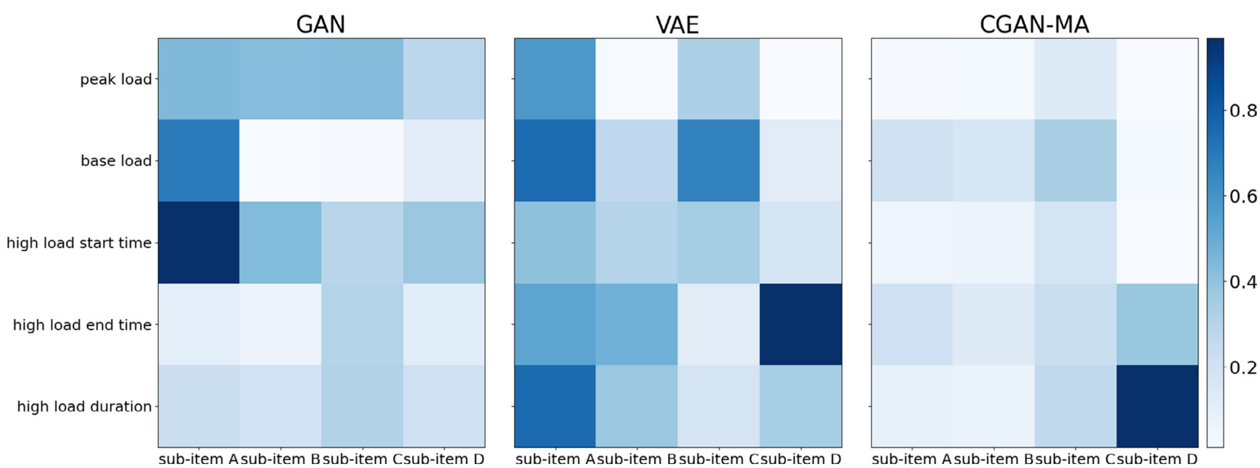


Fig. 9. The KL divergence of key parameters of four sub-item load profiles generated by GAN, VAE, CGAN-MA for office building.

Table 3
Variability of sub-item load profiles generated by different models.

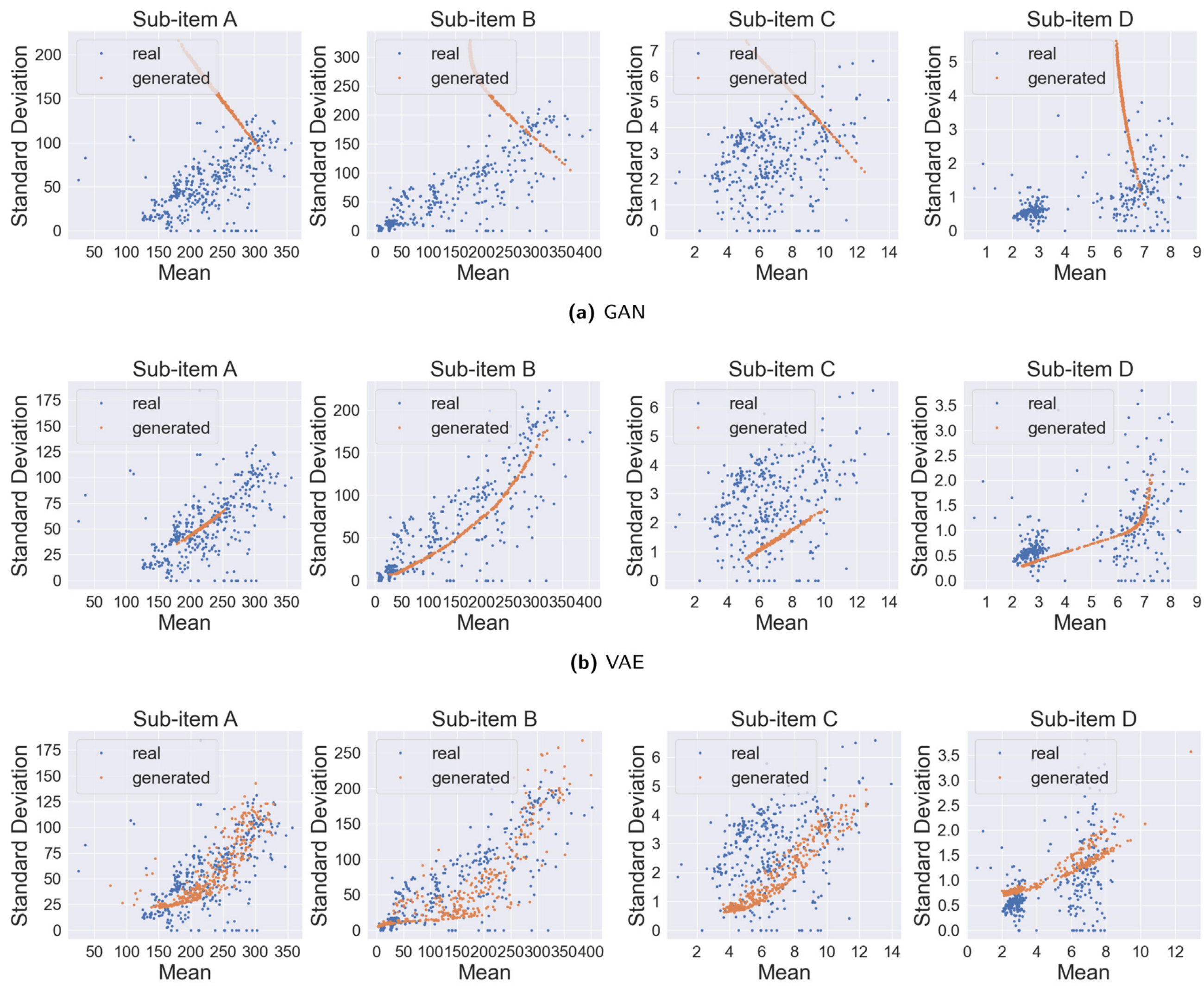
Model	A	B	C	D
Baseline	4.06	10.82	6.92	6.04
GAN	1.90	5.67	2.80	2.62
VAE	0.75	7.17	1.68	4.52
CGAN-MA	3.20	9.95	4.28	5.51

Table 4
Diversity of sub-item load profiles generated by different models.

Model	A	B	C	D
GAN	6.57	17.70	7.77	7.45
VAE	1.61	1.93	1.55	1.30
CGAN-MA	1.36	1.93	1.81	2.08

and energy consumption prediction and grid operation [47] when real data are difficult to obtain owing to privacy issues, time consumption and so on. The machine learning efficacy is evaluated by the performance of a machine learning model trained on the data set. Because in many scenarios, such as electric load forecasting, we need to use a machine learning model for training, In our study, we fixed the selected machine learning model to compare performance on different data sets. The generated load profile and real load profiles are applied for short-term forecasting which is a common task for smart grids [48] to validate that the generated load profile is machine learning efficacy, which can assist in related research in the future.

According to the previous evaluation results of GAN, VAE and CGAN-MA in terms of similarity, variability and diversity, CGAN-MA performs well with regard to several aspects. Therefore, we will apply the data generated by CGAN-MA and real data to short-term energy consumption prediction. Fig. 11 shows the distribution of the generated sub-item load profile and real sub-item load profile for office building, commercial building and hospital. It can be found that each sub-item load of different buildings pre-

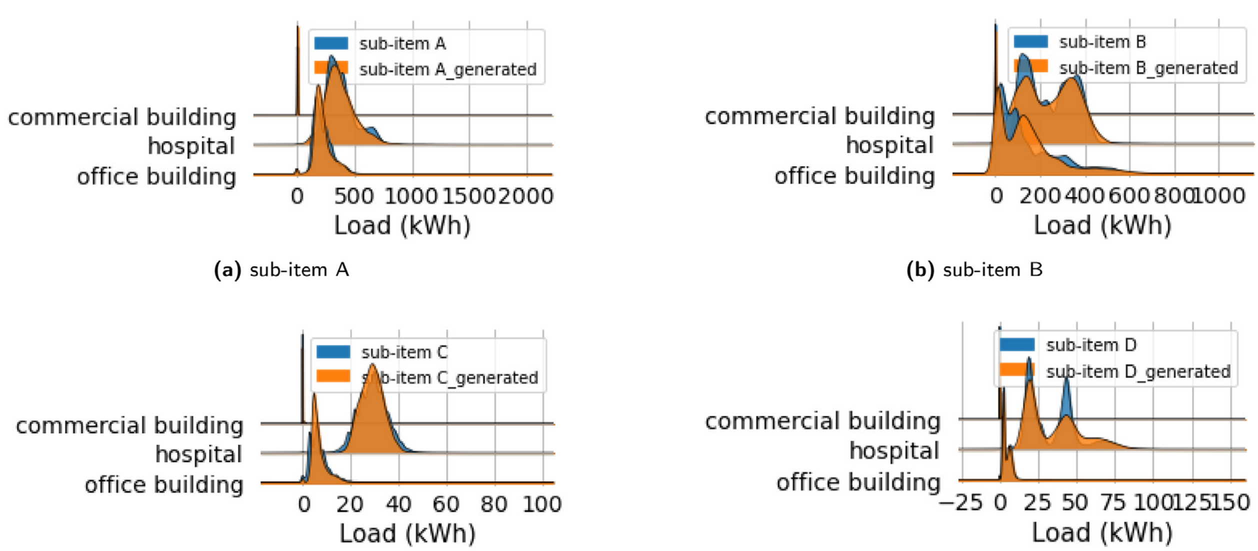


(a) GAN

(b) VAE

(c) CGAN-MA

Fig. 10. Mean-standard deviation distribution of load profile generated by GAN, VAE, CGAN-MA respectively.



(a) sub-item A

(b) sub-item B

(c) sub-item C

(d) sub-item D

Fig. 11. Four sub-item load distribution for different types of buildings.

sents different levels and distribution patterns, and the proposed method can simulate similar levels and patterns of load profiles. Overall, the load level of each sub-item of the hospital is higher than that of office building and commercial building, and the load level of lighting power and air conditioning power is higher than that of impetus power and special power.

We used one month of normalized load data of 2017, including maximum temperature, minimum temperature, weather, wind power, day type, month, and day to train the model and predict the load for a week after the month to test the model. To validate the generation of the generated data, we used a multi-layer perceptron (two fully connected layers and a ReLU activation function) instead of a complex network structure. The multilayer perceptron (MLP) is a feedforward artificial neural network model, which maps multiple input data sets to a single output data set [49]. Re LU activation function is used to increase the nonlinearity of the neural network model. In our study, we used the generated load profile to perform short-term load forecasting tasks based on this model to validate the machine learning efficacy of generated load. First, we sorted the generated daily load profiles according to the generated order to get the generated time-series data. Then, we made predictions on the generated load data and the real load data respectively. The generated data and real data are respectively divided into train set and test set, and then the prediction model was trained on the corresponding train set. The test set is used to test the prediction model. Figs. 12 and 13 show the prediction results of the real total load and generated total load for office building, commercial building and hospital respectively. The three curves in Fig. 12 refer to the results of predicting the generated load profile, the results of predicting the real load profile and the real load profile. It can be found from Fig. 12 that the trend and periodicity of the predicted curve match the ground truth for real load and generated load although there are some extreme values of the predicted curve of the generated total load for commercial buildings and hospitals, which can be attributed to the randomness. Fig. 13 shows the prediction performance on both real total load data and generated total load data considering accuracy indicators, mean absolute error(MAE) and root mean square error (RMSE). As shown in Fig. 13, the model performs well for real data and synthetic data for office building, commercial building, and hospital with MAE and RMSE under 0.06. In addition, the predicted results based on synthetic data and real data are similar in the evaluation of these two indicators. Table 5 lists the MAE and RMSE of the prediction results for office building, commercial building and hospital with regard to real sub-item load profiles and synthetic sub-item load profiles respectively. For the prediction result of the sub-item load profile, the predicted results from synthetic data are similar to those from real data in the estimation of these two indicators; however, for commercial buildings and hospitals, the relative prediction error on synthetic data is slightly larger than that on real data owing to the uncertainty of the generated data to some extent.

4. Result

In this section, we divide each sub-item load profile according to the label. On the one hand, it is similar to the previous verification and analysis of the characteristics of each sub-item load under different labels. On the other hand, it further verifies the load profile under corresponding weather or time conditions, which can be simulated by comparing the generated load possibility distribution with the real load possibility distribution under the same label. We chose three typical labels: day type (time label), weather (weather label), and highest temperature (weather label).

4.1. Generated load profiles under different day types

The four sub-item loads of working days and nonworking days for office building, commercial building and hospital are taken as examples shown in Figs. 14 and 15; They show that for office buildings, the load level on working days is higher than that on non-working days, which is expected for sub-item B(HVAC), and it keeps high value spreading across the middle of the day. For the commercial building, the four sub-item load profiles on working day and non-working day do not have a large difference. For the hospital, the load profile of sub-item A(lighting and socket) on working day is higher than that on non-working days. Moreover, the generated sub-item load distribution under different day types is close to the actual level, indicating that we can obtain each sub-item load on both non-working and working days using this methodology.

4.2. Generated load profiles under different weather

For different weather conditions, we plotted each sub-item load possibility distribution of office buildings, commercial buildings and hospitals under different weather conditions as shown in Fig. 16. We use 0-8 to represent different weather conditions (0: sunny, 1: cloud, 2: overcast, 3: lightning, 4: light rain, 5: model rain, 6: heavy rain, 7: rainstorm, 8: big rain storm). We plot the light and socket load and HVAC load under the same weather conditions (including generated load and real load), and plotted the impetus load and special load together (including generated load and real load). Here, we use a warm color to represent the loads generated under different labels and a cold color to represent the real loads under different labels. It can be found that there are most overlapping areas for the generated and real sub-item load profiles in the overall distribution regardless of the weather conditions. The sub-item loads of office buildings and hospitals change more under

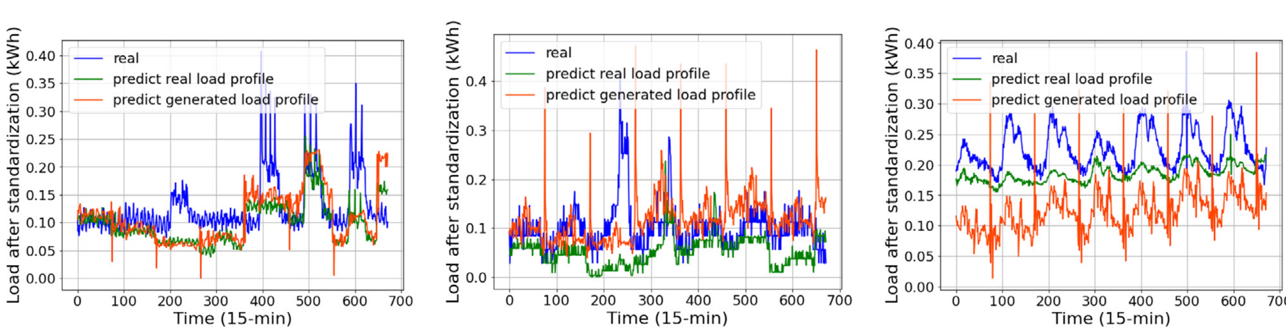


Fig. 12. Prediction on generated load and real load for office building, commercial building and hospital.

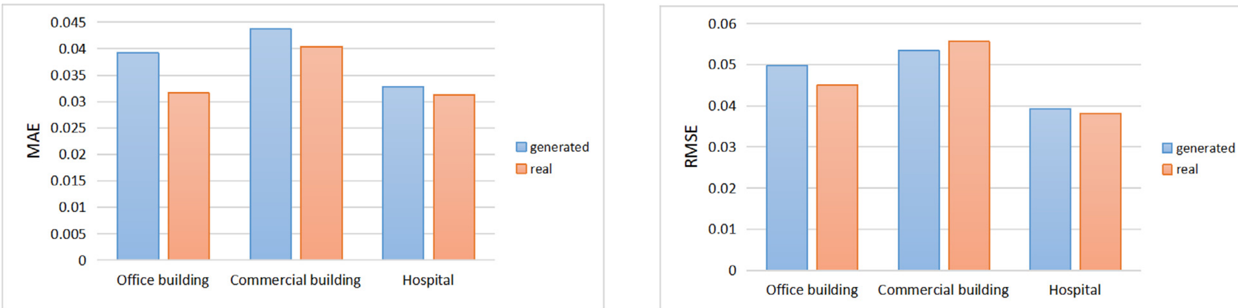


Fig. 13. Prediction performance on generated total load and real total load for office building, commercial building and hospital.

Table 5
Prediction performance on both generated sub-item load profile and real load profile for office building, commercial building and hospital.

sub-item load profile	office building		commercial building		hospital	
	MAE	RMSE	MAE	RMSE	MAE	RMSE
sub-item A(generated)	0.0307	0.0382	0.0915	0.108	0.0365	0.046
sub-item A(real)	0.0328	0.0421	0.0679	0.0848	0.035	0.0427
sub-item B(generated)	0.0357	0.0502	0.0233	0.0433	0.0415	0.0498
sub-item B(real)	0.033	0.0489	0.0159	0.0289	0.019	0.0243
sub-item C(generated)	0.0313	0.0442	0.0626	0.1506	0.0374	0.0474
sub-item C(real)	0.0941	0.1176	0.0541	0.115	0.0453	0.0571
sub-item D(generated)	0.0229	0.0303	0.0737	0.1342	0.0331	0.0423
sub-item D(real)	0.0273	0.0333	0.0459	0.1201	0.0138	0.0213
total load(generated)	0.0392	0.0498	0.0437	0.0534	0.0328	0.0393
total load(real)	0.0316	0.0451	0.0404	0.0556	0.0313	0.0381

different weather conditions than that of commercial buildings. The sub-item loads on rainy days (lightning, light rain, moderate rain, heavy train, rain storm, and large rainstorm) are higher than those under other kinds of weather, especially for HVAC sub-item loads.

4.3. Generated load profiles under different temperature levels

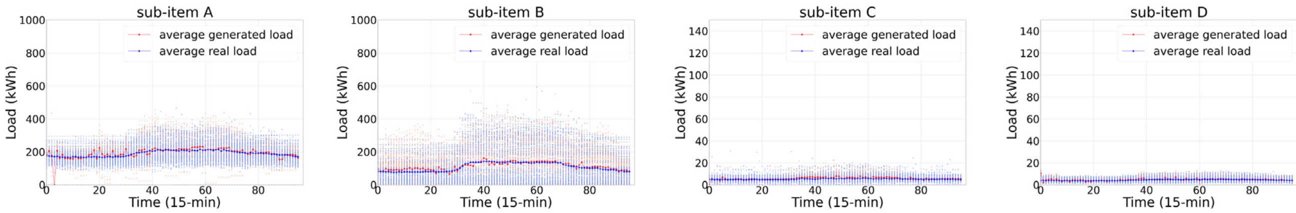
With regard to different temperature levels, we assumed that the lowest temperature and highest temperatures exhibited a similar trend and only used the highest temperature as the label and divided the range of daily highest temperatures across the whole year into three temperature levels (between 20° and 28°, greater than 28°, and less than 20°). Similarly, the lighting and socket load and HVAC load with similar levels are drawn together (including the generated load and real load), and the power load and special power load with similar levels are drawn together (including the generated load and real load). Fig. 17 shows the sub-item load distribution of office buildings, commercial buildings and hospitals under different highest temperature ranges. It can be observed that the light and socket loads and HVAC loads exhibit obviously different distributions at different temperatures, while the distribution of the impetus load and special load at different temperature levels are roughly similar. The entire distribution of light and socket load and HVAC load at a temperature level >28° is higher than those at other temperature levels, which means that the higher the temperature level, the higher the level of light and socket load and HVAC load. This feature is shown in synthetic and real datasets. Therefore, this method can be used to simulate the possible load distribution at the corresponding temperature level.

5. Discussions

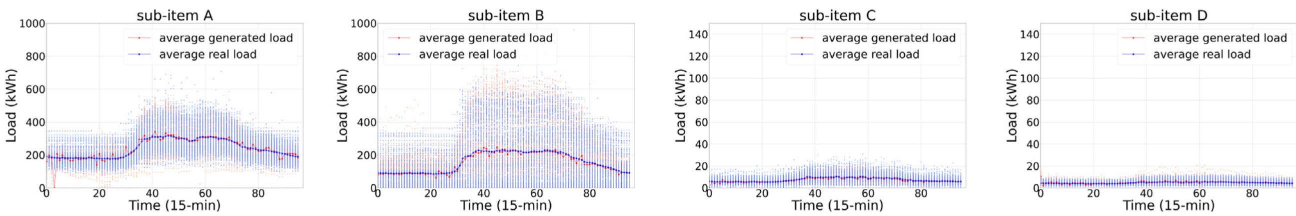
We use different types of buildings (office buildings, commercial buildings, and hospitals) for the case study, rather than considering only one type of building in many related works. In addition, in most of the past studies, the overall load or the overall load data after clustering was considered as the research object, and we used the sub-item load profile (light and socket, HVAC, impetus and special) that was divided based on the purpose of the generation object. The generation effect is improved by considering levels of different sub-items. Using this method, the partial load curve based on external factors (e.g. day type, weather, temperature, and building type) can be simulated.

The CGAN-MA model can well perform in generating the corresponding load profiles under specific conditions. Considering the type of condition, similar to real circumstances, the generated sub-item load on working days is higher than that on non-working days, especially for sub-item A(light and socket consumption) and sub-item B(HVAC consumption) which account for a large part of the total energy consumption. In addition, the generated sub-item A load profiles and generated sub-item B load profiles exhibit a peak period which is similar to the real sub-item A load profile and sub-item B load profiles. Under different weather conditions, the generated load profiles of office buildings and hospitals exhibit more variations than the generated load profile of commercial buildings which may be owing to the low load level. Furthermore, the sub-item B(HVAC) load profile is more sensitive than other sub-item loads under different weather conditions. The load level of sub-item B (HVAC) generated during lighting and rainy days is higher than that during other types of weather conditions. With regard to temperature condition, generated sub-item load is high when the temperature is high, especially obvious for sub-item B(HVAC consumption). It can be concluded that the sub-item B(HVAC) accounts for a large part of the total energy consumption and is the most sensitive sub-item load under different conditions. The load profile generated by CGAN-MA reflects the real circumstances.

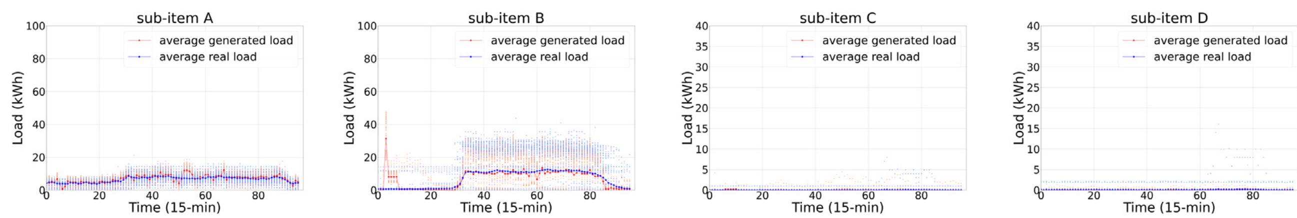
This shows that this method can be applied to simulate the load mode under specific weather and time conditions. The load profile generated under different conditions reflects the electricity



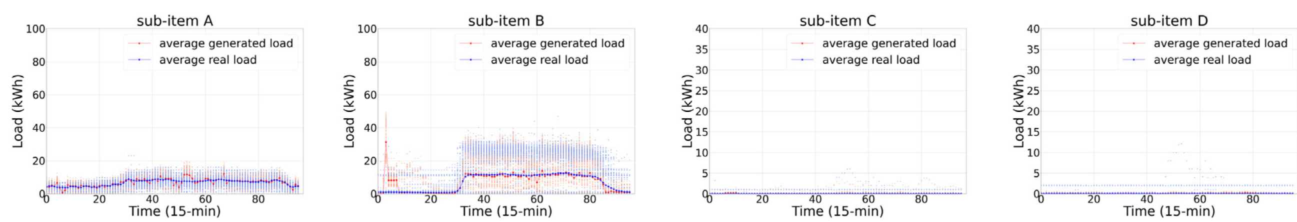
(a) Sub-item load profile(generated and real) on non-working day for office building



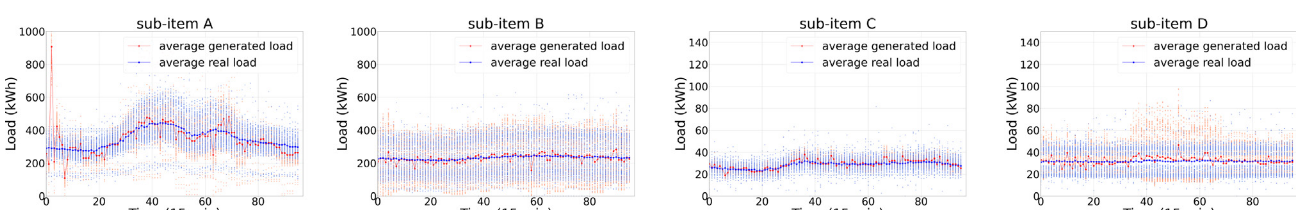
(b) Sub-item load profile(generated and real) on working day for office building



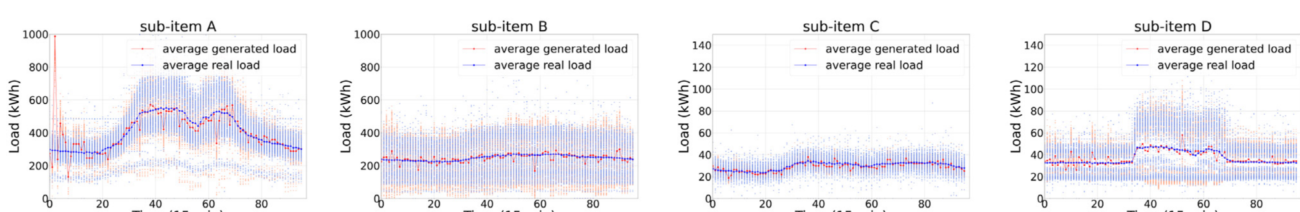
(c) Sub-item load profile(generated and real) on non-working day for commercial building



(d) Sub-item load profile(generated and real) on working day for commercial building



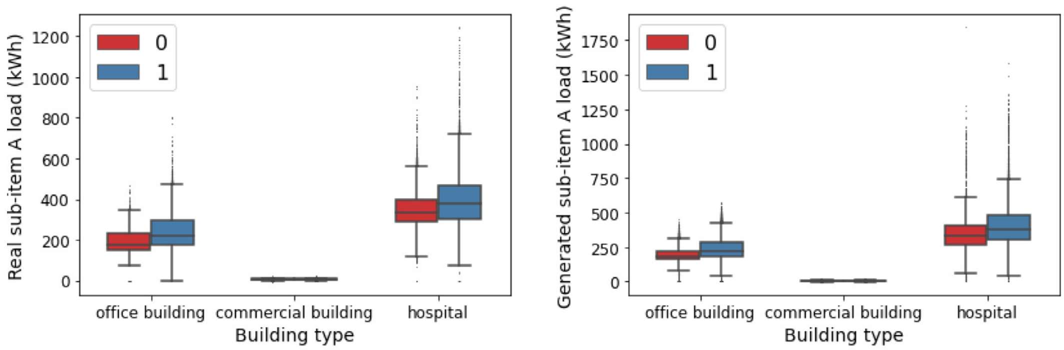
(e) Sub-item load profile(generated and real) on non-working day for hospital



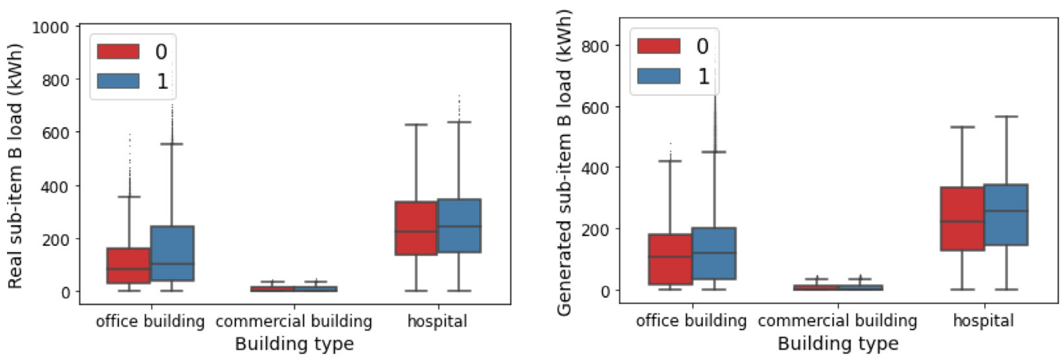
(f) Sub-item load profile(generated and real) on working day for hospital

Fig. 14. Sub-item load profile under different day type.

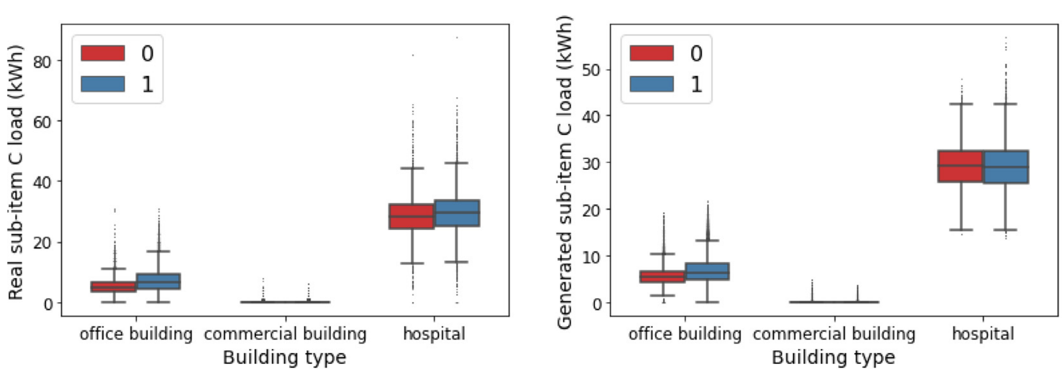
demand under different time conditions and weather conditions. In demand response programs, some incentive measures are taken to adjust the energy consumption behavior. For example, the power company can direct control load by giving customer incentives. If the power company anticipates the demand in advance based on our model, it can better decide how much incentive is



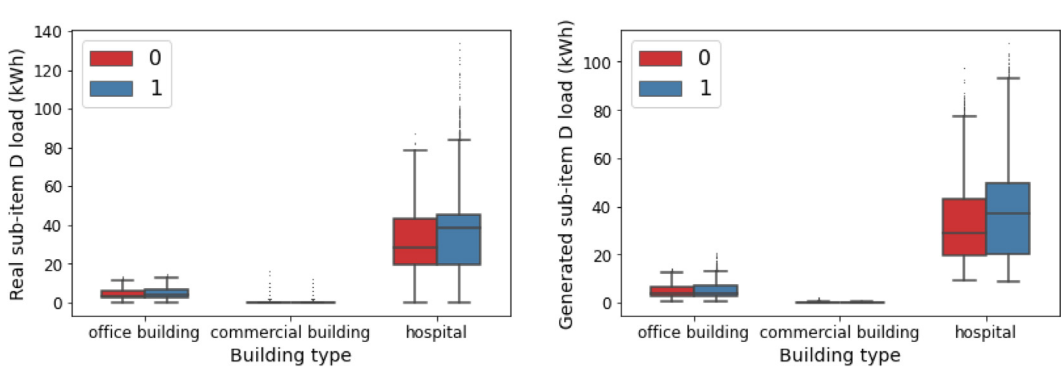
(a) Real and generated sub-item A load profile under nonworking day (0) and working day (1)



(b) Real and generated sub-item B load profile under nonworking day (0) and working day (1)

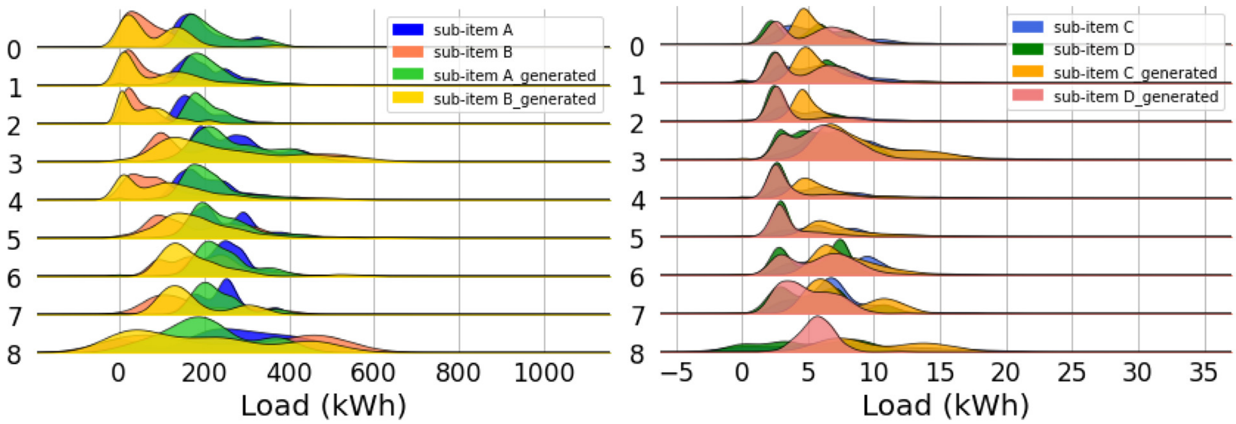


(c) Real and generated sub-item C load profile under nonworking day (0) and working day (1)

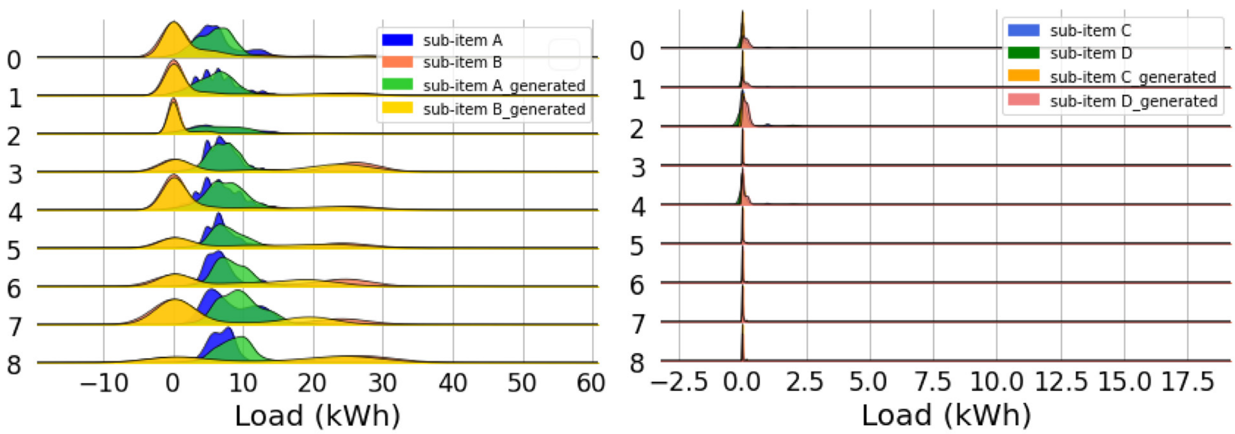


(d) Real and generated sub-item D load profile under nonworking day (0) and working day (1)

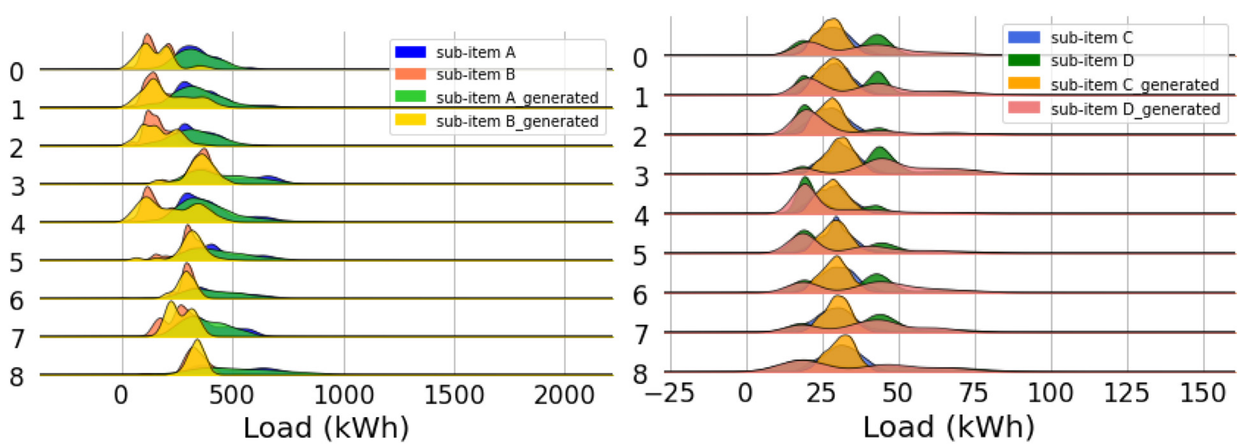
Fig. 15. Real and generated sub-item load profiles under different day types for office building, commercial building and hospital.



(a) Load profile(generated and real) distribution for office building



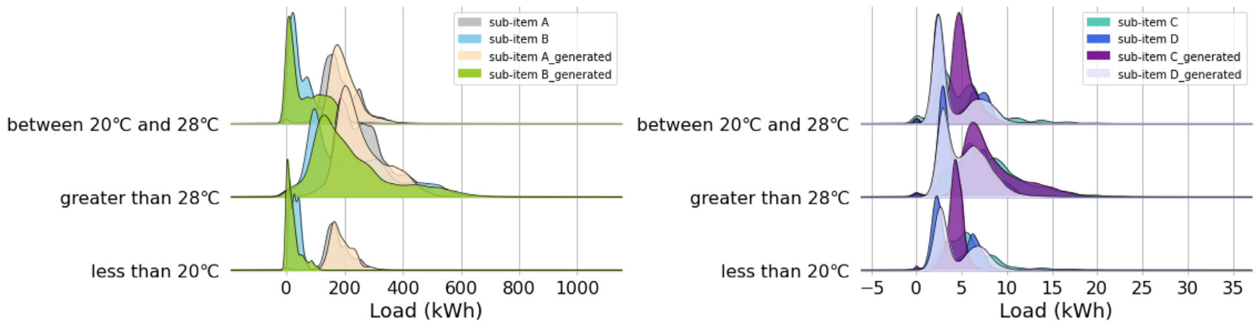
(b) Load profile(generated and real) distribution for commercial building



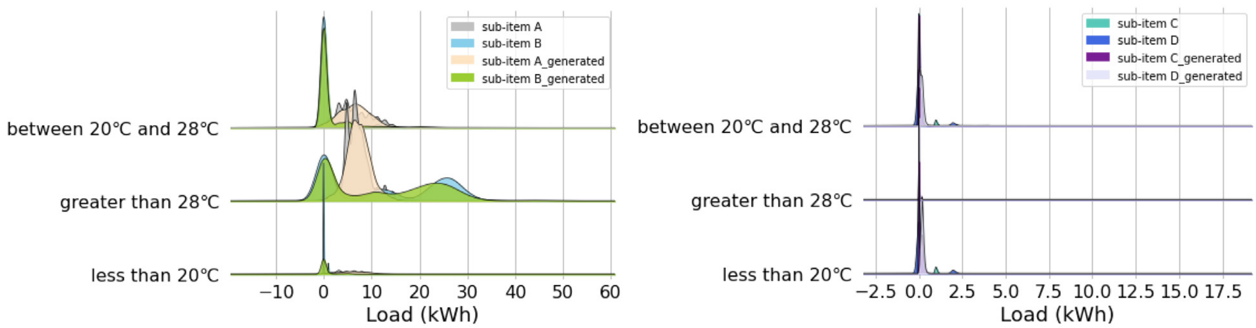
(c) Load profile(generated and real) distribution for hospital

Fig. 16. The generated and real sub-item load distribution under different kinds of weathers for office building,commercial building and hospital.

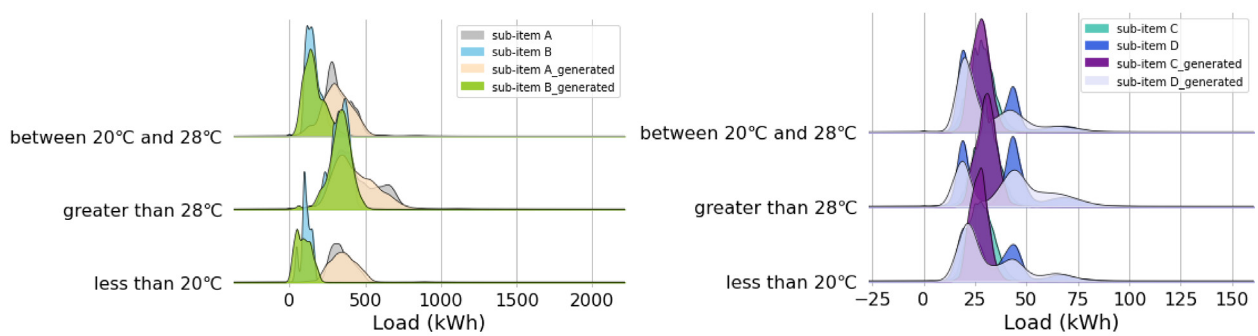
more appropriate for the customer. Different controls shall be made for sub-items with different flexibility. Based on the level and mode of each sub-item load under different conditions, the discriminator can diagnose the abnormal load mode caused by changes in the operation state or other reasons. Owing to these limitations, CGAN-MA is considered as a data-driven method; therefore, the data generated are still based on the existing building data, maintaining statistical characteristics and randomness similar to the original data. In this case, if the building is reconstructed or most tools are updated, the generator trained based on the existing data cannot capture the change in load level caused by the change in the state of the building itself. It is necessary to



(a) Load profile(generated and real) distribution for office building



(b) Load profile(generated and real) distribution for commercial building



(c) Load profile(generated and real) distribution for hospital

Fig. 17. The generated and real sub-item load distribution under different level of temperature for office building,commercial building and hospital.

continue collecting the data of the corresponding buildings. Therefore, collecting more types of building data for research is one of our follow-up studies.

6. Conclusions

In this paper, we propose a novel approach based on Conditional Generative Adversarial Network(CGAN) integrated with the moving average method to generate a sub-item load profile. The proposed CGAN-MA model considers the load shape, weather and time as label inputs to the generator and discriminator to improve the ability to generate sub-item load profiles. By integrating the CGAN with the moving average method, the raw noise was reduced. We validated the generation performance of the sub-item load profile of CGAN-MA by comparing it with the traditional load profile generation method, GAN and VAE, based on three aspects: similarity, variability and diversity. The similarity means that the distribution of key parameters of the generated load profile is similar to that of the real load profile and it is evaluated by the mean difference and KL divergence of key parameters. The variability means that the generated load profiles are different internally, that is, the curves generated each time are different, it is defined as the average of the Euclidean distance between two samples in the generated load profile group. The diversity means that the generated load profile is a batch of new load profiles which are not as same as the real load profiles and it is measured by the average of Euclidean distance between each sample in the generated load profile group and the nearest sample in the real load profile group. The result shows that

1) The proposed model can satisfy requirements in the aspect of diversity with the diversity indicator of four sub-item generated load is 1.36, 1.93, 1.81, 2.08 respectively, and perform best in similarity with the mean difference of key parameters of generated load are most less than 2 and the KL divergence of the key parameters of the generated load are most less than 0.3, which is lower than the load generated by other models. The proposed model exhibits effective performance in variability, with variability indicators of 3.20,9.95,4.28, and 5.51 for four sub-items, which are higher than the load generated by other models. The four sub-item load profiles collected from three different types of buildings including office building, commercial

building, and hospital were employed to perform the analysis using the proposed model.

2) In the case study, the generated and real load profiles are used in the short-term forecasting task, and the MAE and RMSE of the predicted results based on synthetic and real data are less than 0.06, which is similar based on the evaluation of these two indicators. This indicates that the machine learning model can have a performance on generated load profile as good as on real load profile, suggesting the potential of using synthetic data to train machine learning models on various tasks.

3) Furthermore, we validate that the load profile under weather or time conditions can be simulated by comparing the generated load possibility distribution with the real load possibility distribution under the same conditions. The results show that the load generated by CGAN-MA is higher on working days, rainy days and hot days than on non-working days, sunny days and cool days, which correspond to the real circumstances, and sub-item B (HVAC) is the most sensitive sub-item load to different conditions. Therefore, the proposed CGAN-MA can significantly assist load simulation, load forecasting based on external factors and studies related to the operation management of grids.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This study was supported by Basic Research Program of Shenzhen Science and Technology Innovation Committee (WDZC20200819091614001) Shenzhen Science and Technology Program (Grant No. KQTD20170810150821146) and Scientific Research Funds of Tsinghua Shenzhen International Graduate School (Grant No. QD2021007N).

References

[1] C. Ren, S.-J. Cao, Development and application of linear ventilation and temperature models for indoor environmental prediction and HVAC systems control, Sustain. Cities Soc. 51 (2019) 101673, ISSN 2210-6707, doi:10.1016/j.scs.2019.101673..

[2] X. Liu, A. Heller, P.S. Nielsen, CITIESData: a smart city data management framework, Knowl. Inf. Syst. doi:10.1007/s10115-017-1051-3..

[3] Y. Wang, Q. Chen, C. Kang, Q. Xia, M. Luo, Sparse and Redundant Representation-Based Smart Meter Data Compression and Pattern Extraction, IEEE Trans. Power Syst. 32 (3) (2017) 2142–2151, https://doi.org/10.1109/TPWRS.2016.2604389.

[4] S. Joseph, J. Erakkath Abdu, Real-time retail price determination in smart grid from real-time load profiles, Int. Trans. Electr. Energy Syst. 28(3) (2018) e2509, doi:10.1002/etep.2509..

[5] F. Guo, L. Herrera, R. Murawski, E. Inoa, C.-L. Wang, P. Beauchamp, E. Ekici, J. Wang, Comprehensive Real-Time Simulation of the Smart Grid, IEEE Trans. Ind. Appl. 49 (2) (2013) 899–908, https://doi.org/10.1109/TIA.2013.2240642.

[6] Sijie Chen, Chen-Ching Liu, From demand response to transactive energy: state of the art, J. Modern Power Syst. Clean Energy 5 (1) (2017) 10–19.

[7] B.S.G.S.V.S. Vimal, Jesuva Arockiadoss S, Reducing Latency in Smart Manufacturing Service System Using Edge Computing, J. Platform Technol. 9..

[8] K. Zhou, C. Fu, S. Yang, Big data driven smart energy management: From big data to big insights, Renew. Sustain. Energy Rev. 56 (2016) 215–225, ISSN 1364-0321, doi:10.1016/j.rser.2015.11.050..

[9] K. Amasyali, N.M. El-Gohary, A review of data-driven building energy consumption prediction studies, Renew. Sustain. Energy Rev. 81 (2018) 1192–1205, ISSN 1364-0321, doi:10.1016/j.rser.2017.04.095..

[10] Y. Han, B.W. Hong, Deep learning based on Fourier convolutional neural network incorporating random kernels, Electronics 10. doi:10.3390/electronics10162004..

[11] M.U.F.U.M.K.R.S.U.A.H.E. e. Hussain, T., S.W. Baik, Smart and intelligent energy monitoring systems: A comprehensive literature survey and future research guidelines, Int. J. Energy Res. 45. doi:10.1002/er.6093..

[12] R.Y.C.K. Md Ibrahim Khalil, C. Seo, Challenges and Opportunities of Big Data, J. Platform Technol. 8..

[13] M. Bourdeau, X. qiang Zhai, E. Nefzaoui, X. Guo, P. Chatellier, Modeling and forecasting building energy consumption: A review of data-driven techniques, Sustain. Cities Soc. 48 (2019) 101533, ISSN 2210-6707. doi:10.1016/j.scs.2019.101533..

[14] J. Hu, A.V. Vasilakos, Energy Big Data Analytics and Security: Challenges and Opportunities, IEEE Trans. Smart Grid 7 (5) (2016) 2423–2436, https://doi.org/10.1109/TSG.2016.2563461.

[15] A. Molina-Markham, P. Shenoy, K. Fu, E. Cecchet, D. Irwin, Private Memoirs of a Smart Meter, in: Proceedings of the 2nd ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Building, BuildSys '10, Association for Computing Machinery, New York, NY, USA, 61–66, ISBN 9781450304580, 2010, doi:10.1145/1878431.1878446..

[16] Z. Wang, T. Hong, Generating realistic building electrical load profiles through the Generative Adversarial Network (GAN), Energy Build. 224 (2020) 110299, ISSN 0378-7788, doi:10.1016/j.enbuild.2020.110299..

[17] W. Wang, T. Hong, N. Li, R.Q. Wang, J. Chen, Linking energy-cyber-physical systems with occupancy prediction and interpretation through WiFi probe-based ensemble classification, Appl. Energy 236 (FEB.15) (2019) 55–69, https://doi.org/10.1016/j.apenergy.2018.11.079.

[18] Z. Wang, T. Hong, M.A. Piette, Data fusion in predicting internal heat gains for office buildings through a deep learning approach, Appl. Energy 240 (15) (2019) 386–398, https://doi.org/10.1016/j.apenergy.2019.02.066.

[19] E. Nashrullah, A. Halim, Performance Evaluation of Median Filter In Modeling Household Electric Load Profile, in: International Conference on Nano Electronics Research and Education;Takayanagi Kenjiro Memorial Symposium, 2018.

[20] Q. Wei, X.L. Wei, Electric Power Load Profile Estimation Applying Kernel Independent Component Analysis, Zhongguo Dianji Gongcheng Xuebao/Proceedings of the Chinese Society of, Electr. Eng. 28 (16) (2008) 103–107.

[21] N.H. Tehrani, P. Wang, Probabilistic estimation of plug-in electric vehicles charging load profile, Electr. Power Syst. Res. 124 (jul.) (2015) 133–143. doi:10.1016/j.epsr.2015.03.010..

[22] Z. Pan, J. Wang, W. Liao, H. Chen, D. Yuan, W. Zhu, X. Fang, Z. Zhu, Data-Driven EV Load Profiles Generation Using a Variational Auto-Encoder, Energies 12 (5), ISSN 1996–1073, doi:10.3390/en12050849..

[23] L. Ge, W. Liao, S. Wang, B. Bak-Jensen, J.R. Pillai, Modeling Daily Load Profiles of Distribution Network for Scenario Generation Using Flow-Based Generative Network, IEEE Access 8 (2020) 77587–77597, https://doi.org/10.1109/ACCESS.2020.2989350.

[24] A. Creswell, T. White, V. Dumoulin, K. Arulkumaran, B. Sengupta, A.A. Bharath, Generative Adversarial Networks: An Overview, IEEE Signal Process. Mag. 35 (1) (2018) 53–65, https://doi.org/10.1109/MSP.2017.2765202.

[25] L. Abady, M. Barni, A. Garzelli, B. Tondi, GAN generation of synthetic multispectral satellite images, Image and Signal Processing for Remote Sensing XXVI (2020), https://doi.org/10.1117/12.2575765.

[26] Y. Gu, Q. Chen, K. Liu, L. Xie, C. Kang, GAN-based Model for Residential Load Generation Considering Typical Consumption Patterns, in: 2019 IEEE Power Energy Society Innovative Smart Grid Technologies Conference (ISGT), 1–5, 2019, doi:10.1109/ISGT.2019.8791575..

[27] J. Moon, S. Jung, S. Park, E. Hwang, Conditional Tabular GAN-Based Two-Stage Data Generation Scheme for Short-Term Load Forecasting, IEEE Access 8 (2020) 205327–205339, https://doi.org/10.1109/ACCESS.2020.3037063.

[28] Y. Chen, U. Berardi, F. Zhang, Day-ahead prediction of hourly subentry energy consumption in the building sector using pattern recognition algorithms, Energy (2020), https://doi.org/10.1016/j.energy.2020.118530 118530.

[29] Y. Chen, H. Tan, Short-term prediction of electric demand in building sector via hybrid support vector regression, Appl. Energy (2017) S0306261917303082. doi:10.1016/j.apenergy.2017.03.070..

[30] A. Dhar, T.A. Reddy, D.E. Claridge, Generalization of the Fourier series approach to model hourly energy use in commercial buildings, J. Sol. Energy Eng. 121 (1) (1999) 54–62, https://doi.org/10.1115/1.2888143.

[31] E. Jianwei, J. Ye, L. He, H. Jin, Energy price prediction based on independent component analysis and gated recurrent unit neural network, Energy 189 (2019) 116278–. doi:10.1016/j.energy.2019.116278..

[32] H. Guo, D. Hou, S. Du, L. Zhao, N. Yan, A driving pattern recognition-based energy management for plug-in hybrid electric bus to counter the noise of stochastic vehicle mass, Energy 198 (2020), https://doi.org/10.1016/j.energy.2020.117289 117289.

[33] W. Yi, Q. Chen, C. Kang, Q. Xia, Clustering of Electricity Consumption Behavior Dynamics Toward Big Data Applications, IEEE Trans. Smart Grid 7 (5) (2016) 2437–2447, https://doi.org/10.1109/TSG.2016.2548565.

[34] H. Li, Z. Wang, T. Hong, A. Parker, M. Neukomm, Characterizing patterns and variability of building electric load profiles in time and frequency domains, Appl. Energy 291 (2021) 116721, ISSN 0306-2619. doi:10.1016/j.apenergy.2021.116721..

[35] X. Luo, T. Hong, Y. Chen, M.A. Piette, Electric load shape benchmarking for small- and medium-sized commercial buildings, Appl. Energy 204 (2017) 715–725, ISSN 0306-2619, doi:10.1016/j.apenergy.2017.07.108..

[36] P. Price, Methods for Analyzing Electric Load Shape and its Variability, Lawrence Berkeley National Laboratory doi:10.2172/985909..

[37] S.R. Kuppannagari, R. Kannan, V.K. Prasanna, Optimal Net-Load Balancing in Smart Grids with High PV Penetration, BuildSys '17, Association for Computing Machinery, New York, NY, USA, ISBN 9781450355445, doi:10.1145/3137133.3137145..

[38] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative adversarial nets, Advances in neural information processing systems 27. URL: https://proceedings.neurips.cc/paper/2014/file/5ca3e9b122f61f8f06494c97b1afccf3-Paper.pdf..

[39] M. Mirza, S. Osindero, Conditional generative adversarial nets, arXiv preprint arXiv:1411.1784 doi:10.48550/arXiv.1411.1784..

[40] Y. Kim, Convolutional Neural Networks for Sentence Classification, Eprint Arxiv doi:10.3115/v1/D14-1181..

[41] J.S. Hunter, The exponentially weighted moving average, J. Qual. Technol. 18 (4) (1986) 203–210, https://doi.org/10.1002/9780470400531.eorms0314.

[42] J. Goldberger, S. Gordon, H. Greenspan, et al., An Efficient Image Similarity Measure Based on Approximations of KL-Divergence Between Two Gaussian Mixtures., in: ICCV, vol. 3, 487–493, 2003, doi:10.1109/ICCV.2003.1238387..

[43] K. Shmelkov, C. Schmid, K. Alahari, How good is my GAN?, in: Proceedings of the European Conference on Computer Vision (ECCV), 2018, https://doiorg/10.48550/arXiv.1807.09499.

[44] M. Quintana, S. Schiavon, K.W. Tham, C. Miller, Balancing thermal comfort datasets: We GAN, but should we?, in: Proceedings of the 7th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation, 120–129, 2020, doi:10.1145/3408308.3427612..

[45] J. Wu, Z. Lian, Z. Zheng, H. Zhang, A method to evaluate building energy consumption based on energy use index of different functional sectors, Sustain. Cities Soc. 53 (2020), https://doi.org/10.1016/j.scs.2019.101893 101893.

[46] Z. Wang, T. Hong, Generating realistic building electrical load profiles through the Generative Adversarial Network, Energy Build. 224. doi:10.1016/j.enbuild.2020.110299..

[47] F. Guo, L. Herrera, R. Murawski, E. Inoa, C.-L. Wang, P. Beauchamp, E. Ekici, J. Wang, Comprehensive real-time simulation of the smart grid, IEEE Trans. Ind. Appl. 49 (2) (2013) 899–908, https://doi.org/10.1109/TIA.2013.2240642.

[48] C. Zhang, S.R. Kuppannagari, R. Kannan, V.K. Prasanna, Generative adversarial network for synthetic time series data generation in smart grids, in: 2018 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm), IEEE, 1–6, 2018, doi:10.1109/SmartGridComm.2018.8587464..

[49] N. Aerrabotu, G.A. Tagliarini, E.W. Page, S.K. Rogers, Ensemble encoding for time series forecasting with MLP networks, in: Proceedings of SPIE – The International Society for Optical Engineering 3077, 1997, pp. 84–89, https://doi.org/10.1117/12.271468.