Evaluating Performance of Different Generative Adversarial Networks for Large-Scale Building Power Demand Prediction

Yunyang Ye[a,*], Matthew Strong[b], Yingli Lou[b], Cary A. Faulkner[b], Wangda Zuo[c,d], Satish Upadhyaya[b]

[a]*Pacific Northwest National Laboratory, Richland, WA 99354, USA*
[b]*University of Colorado Boulder, Boulder, CO 80309, USA*
[c]*Pennsylvania State University, University Park, PA 16802 USA*
[d]*National Renewable Energy Laboratory, Golden, Colorado, CO 80303 USA*

Abstract

As an unsupervised-learning data-driven model, Generative Adversarial Networks (GANs) have recently attracted a lot of attention for various applications. There is potential to apply GANs for large-scale building power demand prediction, which is needed for power grid operation. However, there are many GAN variations and it is unclear which GAN is suitable for this application. To answer this question, this paper identifies five promising GANs (Original GAN, cGAN, SGAN, InfoGAN, and ACGAN) and evaluates their performance for predicting building power demand at a large scale. Physics-based building energy models are developed to generate training and reference data. A new evaluation indicator that combines accuracy and reproducibility is proposed to evaluate the performance of different GANs in predicting building power demand. The results show that SGAN and InfoGAN are not suitable because they cannot control the number of generated building samples for different building types. The prediction performance among the Original GAN, cGAN, and ACGAN can vary depending on training sample sizes and number of building types. If the training sample size is sufficiently large, Original GAN and cGAN can predict building power demand more accurately than ACGAN with the same number of samples. If training samples are limited, Original GAN provides better accuracy than

*Corresponding author.
Email address: yunyang.ye@pnnl.gov (Yunyang Ye)

cGAN and ACGAN. When the number of building types increase, the prediction accuracy increases for cGAN, decreases for ACGAN, and remains the same for Original GAN. As a result, cGAN and Original GAN are recommended for large-scale building power demand prediction.

Keywords: Generative adversarial networks, data-driven model, building, power demand prediction

---

1. Introduction

Rapid deployment of electrification to address goals of reducing carbon emissions will lead to significant challenges to the power grid. Buildings show great potential for addressing these challenges, since they consume approximately 75% of electricity in the U.S [1] and can be flexible demand-side resources to support the optimal and stable operation of the power grid [2, 3]. To investigate these new challenges, the first step is to precisely predict building power demand at a large-scale. Physics-based simulation methods are often used for large-scale building power prediction. These methods rely on building energy simulation tools, such as DOE-2 [4] and EnergyPlus^{TM} [5]. Urban scale energy modeling and simulation platforms, such as Virtual EPB [6], CitySim [7], and URBANopt [8], were all developed based on building energy modeling tools. While physics-based methods have been proven to be useful for predicting building power demand at a large scale, they have some limitations. This includes requiring expert knowledge and complex workflows to create and calibrate models [9, 10], as well as significant computational time to run hundreds of thousands of detailed building energy simulations for large-scale building power demand prediction.

Data-driven methods are another option for predicting building power demand, which can address the limitations of physics-based methods. These methods utilize historical data to predict future power demand [11, 12, 13]. Recently, some data-driven urban-scale modeling frameworks were developed to predict large-scale building power demand [14, 15, 16, 17]. For example, Yang et al. [17] developed a data-driven urban-scale modeling framework for energy benchmarking of buildings based on recursive partitioning and stochastic frontier analysis. Over 10,000 buildings were involved in this study. The results indicated that this data-driven framework is robust and able to generate accurate energy benchmarking. However, one common limitation of data-driven methods is the requirement of sufficient historical data to provide accurate predictions. Generative adversarial networks (GANs) have become a popular unsupervised-learning data-driven method for various applications that address this limitation. A GAN model is not only able to capture the features of provided data, but also able to keep the individualities of each sample [18]. Torres et al. [19] and Tian et al. [20] pointed out that a GAN model is potentially able to accurately predict time-series building power demand with limited information. This makes GANs a potentially good option for fast and accurate large-scale building power demand prediction with a limited need for historical data.

However, the original GAN model has limitations for more complex studies. For example, the original GAN model is designed to generate samples for a single class, while large-scale power demand prediction involves multiple classes (e.g., building types). To solve this issue, some researchers conducted clustering to classify building samples and then, trained a GAN model for each class [21]. Recently, hundreds of variations of GANs have been proposed to address the limitations of the original GAN model [22], including models with the ability to generate samples for multiple classes. Besides first-clustering-then-GAN approach, these variations of GANs are also potential candidates for multiple-class study. It remains unclear, however, which of these GANs are suitable for building power demand prediction at a large scale.

To answer this question, this paper identifies five promising GANs (Original GAN, cGAN, SGAN, InfoGAN, and ACGAN) and evaluates their performance for predicting building power demand at a large-scale. Physics-based building energy models are developed to generate training and reference data. In addition, a new evaluation indicator is proposed to evaluate the performance of different GANs to predict building power demand. Our scientific contributions include: 1) systematically evaluating the performance of different GAN models for large-scale building power demand prediction, 2) developing a new evaluation indicator to combine the impacts of prediction accuracy and reproducibility, and 3) identifying two GAN models that are not suitable for large-scale building power demand prediction with current methods, as well as one improved GAN model that performs well for several building types and limited training samples.

The rest of this paper is organized as follows: Section 2 selects candidate GANs used in this study; Section 3 proposes the evaluation strategy to determine the suitable GANs for large-scale building power demand prediction; Section 4 prepares training data and reference data using physics-based building energy models; Section 5 displays the evaluation results and conducts discussion based on the results; and finally, Section 6 provides conclusions.

2. GANs Used in the Study

This section introduces variations of GAN and selects a few GANs potentially suitable for predicting large scale building power demand to be further investigated in this study. Subsection 2.1 reviews different GANs and selects five GANs: Original GAN, cGAN, SGAN, InfoGAN, and ACGAN; Subsection 2.2 introduces the Original GAN; Subsection 2.3 introduces the other four selected variations of GAN created based on the Original GAN.

2.1. Review of GANs

Existing studies utilized different GANs to conduct predictions for multiple classes. Tian et al. [20] and Wang and Hong [23] predicted building power demand by using Original GAN models. When there are multiple classes, one Original GAN model is used to generate new samples for each class. Based on the Original GAN model, new GAN variations have been developed and used for various applications. For example, by incorporating semi-supervised learning technology [24], some studies used SGAN [25] for applications, such as fault detection and diagnosis for building heating, ventilation, and air conditioning (HVAC) systems [26, 27]. Wu et al. [28] used InfoGAN [29] and cGAN [30] for energy consumption forecasting in large commercial buildings. Faulkner et al. [31] predicted indoor airflow by using cGAN. In addition, Tang et al. [32] proposed a smart approach to programmatic data augmentation method by using ACGAN [33]. Those previous studies demonstrate that these GAN models are qualified to conduct time-series data prediction with multiple classes, which is the goal of large-scale building power demand prediction. Due to similarities in our application to previous applications, the Original GAN, cGAN, SGAN, InfoGAN, and ACGAN used in those studies are identified as promising GANs for this study. The following two subsections will detail these five GANs.

2.2. Original GAN

The Original GAN contains two neural networks that compete against each other in a game [18]. The two neural networks are referred to as the generator and discriminator, respectively. The objective function of the Original GAN is expressed as:

$$\min_{G} \max_{D} V(D,G) = \mathbb{E}_{x\sim p_{data}(x)}\left[logD(x)\right] + \mathbb{E}_{z\sim p_{z}(z)}\left[logD\left(1 - D\left(G\left(z\right)\right)\right)\right], \tag{1}$$

where D is the discriminator; G is the generator; $\mathbb{E}$ is the expectation; $x \sim p_{data}(x)$ is the real samples, following the probability, p_{data}. The $z \sim p_z(z)$ is the input noise vector for the generator and z follows the probability, $p_z(z)$. The objectives of the Original GAN training are (1) to train D to maximize the probability of assigning the correct labels to both real samples and generated samples, and (2) to train G to minimize the probability of D to distinguish between real samples and generated samples.

Based on samples with common characteristics, the Original GAN shows excellent performance to generate similar samples. However, power demands in different buildings can significantly vary due to different building characteristics and operation patterns. To predict building power demand at a large scale, a large number of different buildings need to be studied. Thus, it is necessary to classify the studied buildings into multiple groups. Each group of buildings share similarities, such as similar building characteristics and operation patterns. This paper groups buildings by using building type. Figure 1 shows the method to train Original GAN models for multiple building types. When the Original GAN is used to predict building power demand at a large scale, one Original GAN model is required for each building type.

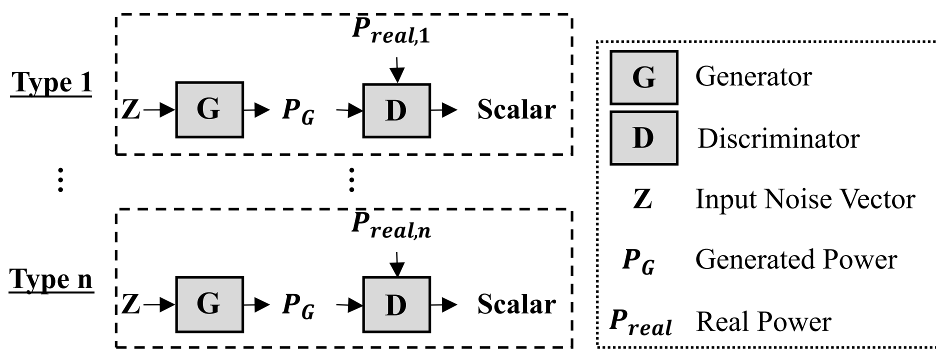

Figure 1: Model training for multiple building types using the Original GAN.

2.3. Four Variations of GAN

One Original GAN model can only generate data for one class. When a study is complex and contains many classes (e.g., different building types), significant effort is required to train multiple Original GAN models. To address this issue, many variations of GAN have been developed based on the Original GAN model. This paper selects four GANs among these variations of GAN. Variations of GAN (e.g., Wasserstein GAN [34]) that are not designed to generate samples for multiple types are not considered in this paper. Table 1 summarizes the differences of the four studied GANs from the Original GAN.

cGAN was proposed by [30]. The Original GAN is extended to a conditional model that both the generator and discriminator are conditioned on some extra information. The extended model is cGAN. The extra information is named as auxiliary information. The auxiliary information can be a label corresponding to a specific group of samples, for example.

SGAN was proposed by [25]. SGAN incorporates semi-supervised learning into the Original GAN. Semi-supervised learning [24, 35] provides the possibility to generate data with partial labeled data. Thus, a semi-supervised learning model is able to conduct prediction for multiple categories of samples.

InfoGAN was proposed by [29]. InfoGAN does not use labeled data, but instead decomposes the input noise vector into two parts: source of incompressible noise and latent code that provides salient features about the data distribution. Since the features of different groups of samples should significantly vary, it is possible that the latent code provides the information about the group label.

ACGAN was proposed by [33]. Similar to cGAN, ACGAN also uses auxiliary information as an input for the generator. Unlike cGAN, its discriminator does not receive auxiliary information, and instead determines a probability over the group labels to differentiate between real samples and generated samples.

Figure 2 shows the methods to train GAN models for multiple building types using these four GANs. In each case, only one GAN model needs to be trained. To train a cGAN model (Figure 2(a)), building type is used as the auxiliary information and becomes an input for the generator. To train a SGAN model (Figure 2(b)), the building type is used as one input of the supervised discriminator. For InfoGAN (Figure 2(c)), the physical meaning

Table 1: Changes of four GANs from the Original GAN.

No.	Name	Changes from the Original GAN
1	cGAN [30]	Extends the Original GAN to a conditional model by providing auxiliary information (e.g., labeled data).
2	SGAN [25]	Incorporates semi-supervised learning [35, 24] into the GAN.
3	InfoGAN [29]	Decomposes the input noise vector into two parts: (1) source of incompressible noise and (2) latent code, which targets the salient structured semantic features of the data distribution.
4	ACGAN [33]	Provides a label for each generated samples, and its objective function evaluates both generated samples and their labels.

of the latent code is not defined before the model is trained and the latent code may not represent building type. Thus, the features of the generated building power demand are used to determine the building type after the InfoGAN model is trained, which will be introduced in Subsection 5.1. Similar to cGAN, an ACGAN model (Figure 2(d)) also uses building type as auxiliary information to be used as an input for the generator. The main difference between cGAN and ACGAN is in the discriminator model: the cGAN's discriminator is provided with both the data and class label (building type) as input, whereas the ACGAN's discriminator is only provided with the data as an input, which means that the discriminator must also predict the class label of the sequence.

Figure 3 displays the workflow of predicting the large-scale building power demand using GAN in this study. First, Physics-based building models are used to generate reference power demand set and these power demands are real samples in this study. The physics-based building models will be detailed in Section 4. From the reference power demand set, a small set of data are randomly selected as the training set, which is used to train GAN models. Then, the trained GAN models are used to generate a large set of building power demand data. The performance of these trained GAN models is evaluated by comparing these predicted power demand data with the reference power demand set.

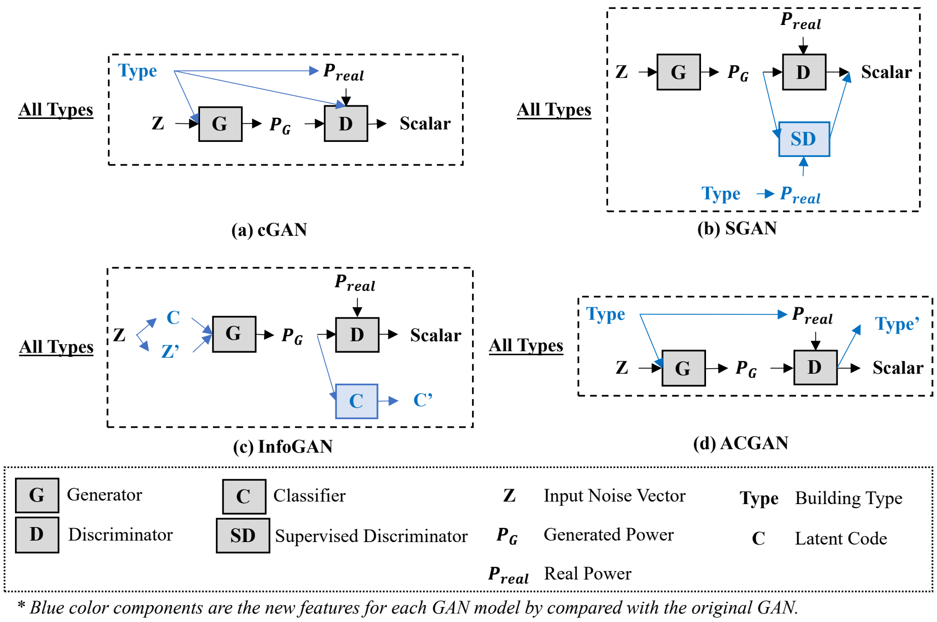

Figure 2: Model training for multiple building types using (a) cGAN, (b) SGAN, (c) InfoGAN, and (d) ACGAN.

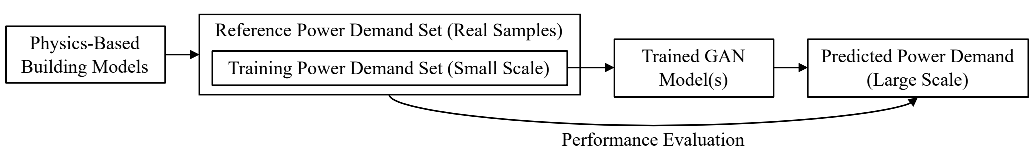

Figure 3: Workflow of predicting the large-scale building power demand using GAN in this study.

3. Evaluation Strategy

This section introduces a strategy to evaluate the five studied GANs for large-scale building power demand prediction in two aspects: suitability and sensitivity. Subsection 3.1 investigates the evaluation strategy for the model suitability by setting a series of application requirements. Subsection 3.2 introduces the evaluation strategy for the sensitivity analysis. Subsection 3.2.1 introduces the evaluation indicator used in the sensitivity analysis while Subsection 3.2.2 introduces the sensitive aspects, including the impact of training size and number of building types to the performance of large-scale building power demand prediction. Based on the model suitability and sensitivity analysis, we will recommend the GAN types for large-scale building power demand prediction among these five studied types. To avoid any bias during the evaluation process, we keep the same settings for generators and discriminators in all studied GAN types.

3.1. Model Suitability

This subsection provides the evaluation strategy about the suitability for each studied GAN. There are two criteria to select GAN types: (1) A GAN model must have the capability of generating the required number of building samples for each building type. Different quantities of building samples are assumed for the studied building types. It is required that a trained GAN model can generate the required quantity of samples for each building type easily. (2) The generated power demand data needs to capture the main features of real data. For example, most office buildings are occupied during the daytime and unoccupied during the nighttime. If generated samples of office buildings tend to have lower power demand during the daytime than nighttime, it means that the GAN model cannot capture this feature of real data. The second criterion is a qualitative evaluation and aims to exclude any GAN type that cannot capture any features of real data. Later, a more detailed quantitative evaluation will be conducted in the model sensitivity, which will be introduced in the next subsection (Subsection 3.2)

If a GAN type does not meet one of the above criteria, it will be excluded from the studied GAN list and we will not study its sensitivity to training size and number of building types introduced in the next subsection.

3.2. Model Sensitivity

This subsection provides the evaluation strategy to analyze the sensitivity for each studied GAN. Subsection 3.2.1 develops a new evaluation indicator to evaluate the sensitivity of each studied GAN; Subsection 3.2.2 introduces the sensitive aspects that this paper studies.

3.2.1. Evaluation Indicator

Some evaluation indicators have already been proposed in existing building power demand prediction studies. For example, Tian et al. [21] used a hybrid physics-based model and generative adversarial network to predict power demand for buildings at a large scale. Mean absolute percentage error (MAPE) was used as an indicator to evaluate the accuracy of the prediction result. Similarly, Bendaoud et al. [36] also used MAPE to evaluate the performance for electricity demand forecasting. In addition, Wang and Hong [23] used difference in the mean, standard deviation, and distribution of key parameters between the generated load profiles and those of the real ones to quantify the accuracy of the load profile generation. The ambient environment and operation in one building contains some uncertainties, which lead to the variability of power demand for this building in different days. When the building power demands are studied at a large scale, their distribution should be more predictable. In general, existing research about evaluating predicted building power only considers accuracy. However, we also need to evaluate if a GAN model is able to predict the distribution stably because the prediction results of GAN model is random. Therefore, it is necessary to consider accuracy as well as reproducibility when predicting building power using GAN. The reproducibility is to evaluate how similar the predicted distribution of large-scale building power demand in multiple runs under the same assumptions.

To consider both, this paper develops a new evaluation indicator. Figure 4 shows the diagram to calculate this evaluation indicator. Step 1 is to evaluate the accuracy of a GAN model; Step 2 is to evaluate the reproducibility of a GAN model; Step 3 combines both accuracy and reproducibility, and generates the final evaluation indicator (EI).

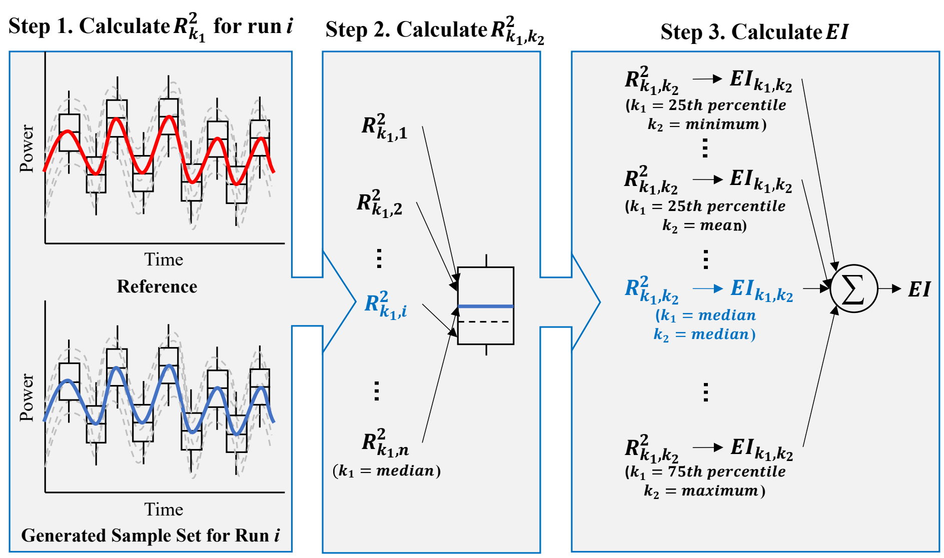

Figure 4: Diagram to calculate evaluation indicators.

In one run, building samples are generated by the trained GAN model. The key parameters of distribution of generated building samples are compared with the real building samples generated by EnergyPlus^{TM}. The key parameters of distribution consist of 25^{th} percentile, median, mean, and 75^{th} percentile values. This study mainly focuses on the distribution of power demand for the majority of buildings. The minimum and maximum values are not considered as key parameters for building power demand since the quantity of building samples with extremely high or low power demand is small. The R^2 is used to reflect the difference of these key parameters between the generated and real samples. The output of the first step is to calculate R^2 of key parameters for each run, and k_1 in Figure 4 is the key parameter index ($k_1 = 25^{th}$ percentile, median, mean, and 75^{th} percentile). $R^2_{k_1}$ is calculated by using the following equation:

$$R^2_{k_1} = 1 - \frac{\sum_{i=1}^{N} \left(y_{k_1,i} - \hat{y}_{k_1,i}\right)^2}{\sum_{i=1}^{N} \left(y_{k_1,i} - \bar{y}_{k_1}\right)^2}, \tag{2}$$

where $y_{k_1,i}$ is a key parameter (k_1) of the generated building power demand at time i ($i = 1, 2, \cdots, N$); $\hat{y}_{k_1,i}$ is the key parameter (k_1) of the real building power demand at time i ($i = 1, 2, \cdots, N$); $\bar{y}_{k_1}$ is the mean of the key parameter (k_1) of the generated building power for all times. A larger value of $R^2_{k_1}$ means that the predicted power demands are more consistent with the real ones.

This paper conducts multiple runs for each type of the selected GAN. In each run, training samples are randomly selected. Then GAN models are trained and new samples are generated using the trained GAN models. Thus, multiple R^2 for each key parameter are obtained.

The output of the step 2 is the distribution of R^2 for each key parameter. The key parameters consist of minimum, 25^{th} percentile, median, mean, 75^{th} percentile, and maximum values. Different from the key parameters for k1, we add minimum and maximum values as key parameters here. This is because the reproducibility also requires evaluating the difference of performance between best and worst runs. k_2 in Figure 4 is the key parameter index of R^2 (k_1 = minimum, 25^{th} percentile, median, mean, 75^{th} percentile, and maximum). It can be expressed as follows:

$$R^2_{k_1,k_2} = f_{k_2}\left(R^2_{k_1,1}, R^2_{k_1,2}, \cdots, R^2_{k_1,n}\right), \tag{3}$$

where k_1 is the key parameter index of building power set in one run; n is the total number of runs; k_2 is the key parameter index of R^2 for k_1 in n runs; f_{k_2} is the formula to calculate the key parameter k_2 for $R^2_{k_1}$ among all runs. A larger value of $R^2_{k_1,k_2}$ means that the predicted power demands are more stable.

The step 3 calculates the evaluation indicator. The following equation transfers $R^2_{k_1,k_2}$ into the evaluation indicator for k_1 and k_2:

$$EI_{k_1,k_2} = \begin{cases} 1 & if\ R^2_{k_1,k_2} \geq 0.95 \\ 0 & otherwise \end{cases}, \tag{4}$$

where EI_{k_1,k_2} is the evaluation indicator for k_1 and k_2.

A $R^2_{k_1,k_2} \geq 0.95$ implies good performance of building power demand performance. Thus, the evaluation indicator gets one credit; otherwise, no credit is gained. The evaluation indicators for all k_1 and k_2 are summed to obtain the final evaluation indicator, which ranges from [0,24]. It is expressed as:

$$EI = \sum_{k_2} \sum_{k_1} EI_{k_1,k_2}. \tag{5}$$

A larger value of EI means that the predicted power demands are more consistent with the real ones and the prediction is more stable. The quantitative analysis considers the impacts of training size, number of epochs, and building type to the evaluation indicator, which are studied when evaluating the performance of different GANs to predict building power demand at a large scale.

3.2.2. Sensitive Aspects

Some factors, such as the training size and number of building types, can potentially affect the performance of large-scale building power demand prediction. In general, increasing training size tends to improve the performance of the GAN models and the generated samples tends to be more accurate. However, collecting a large set of training samples is usually difficult, time consuming, and costly. Thus, it is necessary to conduct sensitivity analysis to evaluate the performance of different GAN models trained using only limited training samples. In addition, involving more building types adds complexity to predict large-scale building power demand. If a type of GAN has similar or even better performance when the number of building types increases, this GAN type is qualified to be used in complex cases.

The above discussion shows that, to comprehensively study the performance of the selected GANs, it is necessary to conduct sensitivity analysis to evaluate the impact of training size and number of building types to the large-scale building power demand prediction. To make an unbiased evaluation, we adopt the same settings of the generator and discriminator for different GANs, which will be detailed in Subsection 5.1. In addition, accuracy and reproducibility are the two main aspects to evaluate the performance of these GANs, which will be quantitatively reflected by using the evaluation indicator introduced in Subsection 3.2.1.

4. Building Power Data Preparation

This section describes the preparation of building power data. This study mainly aims to evaluate the performance of selected GANs for large-scale building power demand. To simplify the study and avoid potential noise in the data, this paper uses physics-based models to generate building power data. We consider different building types, and uncertainties of building characteristics and operation in the building set. The seed models are from DOE's Commercial Prototype Building Models [37] and Residential Prototype Building Models [38]. Based on these seed models, a large set of building samples are generated to evaluate the performance of different GANs for building power demand prediction at a large scale.

First, this paper assumes that all building samples are located in Buffalo, NY (ASHRAE climate zone 5A (cool and moist)). Six building types are selected with different numbers of buildings. Since this paper considers the impact of number of building types, some building types are only used for certain studies. Table 2 displays all selected building types and usages in different studies. Four, five, and six building types are considered in different studies. No.1 to 4 (single-family detached house, small office, standalone retail, and multi-family low-rise apartment building) are selected for all studies. No.5 (medium office) is only selected for the study with six building types. No.6 (primary school) is selected for the studies with both five and six building types. In addition, the numbers of buildings for building types are different in the real world; thus, the number of buildings for the selected building types in this study are also different. The number of buildings for each building type is identified based on the Commercial Buildings Energy Consumption Survey [39], Residential Energy Consumption Survey [40], and engineering judgment.

Table 2: Selected building types and usages in different studies.

No.	Building Type	Number of Buildings	Study of Four Building Types	Study of Five Building Types	Study of Six Building Types
1	Single-family Detached House	4,500	Yes	Yes	Yes
2	Small Office	800	Yes	Yes	Yes
3	Stand-alone Retail	400	Yes	Yes	Yes
4	Multi-family Low-rise Apartment Building	400	Yes	Yes	Yes
5	Medium Office	250	No	No	Yes
6	Primary School	200	No	Yes	Yes

Based on the seed models, 18 input variables are selected and their uncertainties are provided based on [10]. These input variables and their uncertainties are listed in Table 3. The input variables includes building characteristics (e.g., total floor area), building assets (e.g., insulation R-value of roof), and operation patterns (e.g., operation schedule). All input variables are assumed to be uniformly distributed. The ranges of most input variables are ±20% of default values, which are provided by the seed models. The cooling and heating temperature set points are ±1.11°C of default values. Operation hours are extended or reduced by two hours. A large set of building samples are selected and generated among these uncertainties.

After running all simulations using EnergyPlus^{TM}, the power demand for each building sample is generated. The power demand data in August are collected for this study and, in each building sample, the average power demand value in each hour is used. Thus, a series of power demand is prepared for one building sample, which contains 744 values (24 values per day and 31 days in August). All building power demand data are used as reference data. Then, we use a small subset of reference data to train GAN models

Table 3: Selected input variables and their uncertainties.

<table>
<tr><th>No.</th><th>Variable Name</th><th>Range</th></tr>
<tr><td>1</td><td>Total Floor Area[1]</td><td rowspan="15">±20% of Default Value</td></tr>
<tr><td>2</td><td>Floor Height[1]</td></tr>
<tr><td>3</td><td>Aspect Ratio[1]</td></tr>
<tr><td>4</td><td>Window-to-Wall Ratio[2]</td></tr>
<tr><td>5</td><td>Insulation R-value of Roof</td></tr>
<tr><td>6</td><td>Insulation R-value of Exterior Walls</td></tr>
<tr><td>7</td><td>U-factor of Fenestration</td></tr>
<tr><td>8</td><td>SHGC of Window</td></tr>
<tr><td>9</td><td>People Density</td></tr>
<tr><td>10</td><td>Lighting Power Density</td></tr>
<tr><td>11</td><td>Electric Equipment Power Density</td></tr>
<tr><td>12</td><td>Infiltration Rate</td></tr>
<tr><td>13</td><td>Ventilation Rate</td></tr>
<tr><td>14</td><td>Heating Efficiency</td></tr>
<tr><td>15</td><td>Cooling Efficiency</td></tr>
<tr><td>16</td><td>Cooling Temperature Set Point</td><td rowspan="2">±1.11°C of Default Value</td></tr>
<tr><td>17</td><td>Heating Temperature Set Point</td></tr>
<tr><td>18</td><td>Operation Schedule[3]</td><td>±2 Hours of Occupied Time</td></tr>
</table>

[1] Primary school and quick service restaurant do not consider these variables.
[2] Single-family detached house, multi-family low-rise apartment building, primary school and quick service restaurant do not consider this variable.
[3] Single-family detached house and multi-family low-rise apartment building do not consider this variable.

and use trained GAN models to predict the building power demand at a large scale. To train GAN models, the same quantity of the building power demand data for each studied building type are randomly selected. To be used for GAN model training, these training data are normalized using the following equation:

$$P_{norm,i,j,k} = \frac{2 \times P_{i,j,k} - \left(\underset{j,k}{max} \left(P_{i,j,k} \right) + \underset{j,k}{min} \left(P_{i,j,k} \right) \right)}{\underset{j,k}{max} \left(P_{i,j,k} \right) - \underset{j,k}{min} \left(P_{i,j,k} \right)}, \tag{6}$$

where $P_{norm,i,j,k}$ is the normalized building power demand at time k for building j, which belongs to building type i; $P_{i,j,k}$ is the building power demand at time k for building j, which belongs to building type i; $\underset{j,k}{max} \left(P_{i,j,k} \right)$ and $\underset{j,k}{min} \left(P_{i,j,k} \right)$ are the maximum and minimum power demands at all time and for all building samples, which belongs to building type i.

5. Evaluation Results and Discussion

This section analyzes the evaluation results by using the strategy introduced in Section 3. We then discuss the results and recommend the GANs for large-scale building power demand prediction. Subsection 5.1 introduces the model settings for all five studied GANs; Subsection 5.2 analyzes the model suitability for large-scale building power demand prediction; Section 5.3 conducts sensitivity analysis to evaluate the impacts of training size and number of building types to the evaluation performance; Section 5.4 discusses the results.

5.1. GAN Model Settings

Since this paper evaluates the performance of different GANs rather than different model structures for the generator and discriminator, the generator and discriminator both use convolutional neural networks (CNNs) [41] with the same model structure for each GAN type. Figure 5 provides the structures of generators and discriminators in the five studied GANs. A generated sample from each GAN contains 744 values, in order to be evaluated against the ground truth data. The GANs are currently only trained based on building type, but different inputs, such as time, can be included in future work. We have released all scripts and data in `https://github.com/peasant98/GAN-energy-modeling.git`. More detailed information, such as the convolution kernel size, is provided through this link.

Due to the different needs for inputs and outputs in these GANs, the pre-processing and post-processing for both generator and discriminator are

	GAN	cGAN	SGAN	InfoGAN	ACGAN
Generator					
Input	*Z*	*Z* ¦ *Type*	*Z*	*[Z', C]*	*Z* ¦ *Type*
Pre-Processing	Dense LeakyReLU Reshape	Embed ¦ Dense ¦ Dense ¦ LeakyReLU Reshape ¦ Reshape Concat	Dense LeakyReLU Reshape Concat	Dense LeakyReLU Reshape	Embed ¦ Dense ¦ Dense ¦ LeakyReLU Reshape ¦ Reshape Concat
CNN	ConvTranspose LeakyReLU ConvTranspose LeakyReLU Conv	ConvTranspose LeakyReLU ConvTranspose LeakyReLU Conv	ConvTranspose LeakyReLU ConvTranspose LeakyReLU Conv	ConvTranspose LeakyReLU ConvTranspose LeakyReLU Conv	ConvTranspose LeakyReLU ConvTranspose LeakyReLU Conv
Post-Processing	Tanh	Tanh	Tanh	Tanh	Tanh
Output	P_G	P_G	P_G	P_G	P_G
Discriminator					
Input	P_G	P_G ¦ *Type*	P_G	P_G	P_G
Pre-Processing		Embed ¦ Dense ¦ Concat			
CNN	Conv LeakyReLU Conv LeakyReLU Conv LeakyReLU Flatten	Conv LeakyReLU Conv LeakyReLU Conv LeakyReLU Flatten	Conv LeakyReLU Conv LeakyReLU Conv LeakyReLU Flatten	Conv LeakyReLU Conv LeakyReLU Conv LeakyReLU Flatten	Conv LeakyReLU Conv LeakyReLU Conv LeakyReLU Flatten
Post-Processing	Dropout Dense	Dropout Dense	Dropout Dense Lambda ¦ Activation	Dropout Dense ¦ Dense ¦ BatchNormalization ¦ LeakyReLU ¦ Dense	Dropout Dense ¦ Dense
Output	*Scalar*	*Scalar*	*Scalar*	*Scalar* ¦ *C'*	*Scalar* ¦ *Type'*

Figure 5: Structures of generators and discriminators in the five studied GANs.

different, but the structure of the CNN is the same in all five studied GANs. More details can be found in the Github repository, but we follow best practices (which are explained in the Stanford CS231n course [42]) and turn to [43] for generator and discriminator architecture. For clarity, the terminologies in this figure are defined as follows and readers can refer to [18, 30, 33, 41] for more details:

1. Z: This is random noise (the dimension can be selected by a user) that is used as an input for the generator.

2. Type: Building type.

3. Embed: A layer that converts positive integers (indexes) into dense vectors of fixed size.

4. Reshape: A layer that changes the shape (dimensions) of the previous layer to adhere to the model's structure.

5. Dense: A dense layer is the function $\sigma(w^t x + b)$, where $b \in \mathbb{R}^n$, $x \in \mathbb{R}^m$, $w \in \mathbb{R}^{m \times n}$, and σ is a nonlinear activation function.

6. Concat: Concatenates a list of inputs to form one single vector.

7. LeakyReLU: A nonlinear activation function expressed as $\max(0.1x, x)$.

8. Flatten: Flattens the input to be one dimensional. For example, a matrix of dimension $m \times n$ will become a vector of dimension $m \cdot n$

9. Dropout: During training time, we zero out the previous weights in the network with a certain probability, to force other weights to learn.

10. Conv: The main building block of Convolutional Neural Networks that works well on data with a spatial relationship. The convolution layer employs multiple filters that perform convolutions over an image. Please refer to [42] for more information.

11. ConvTranspose: Instead of downsampling layers to reduce amount of information, ConvTranspose, or transpose convolutions, upsample the amount of information. See [44] for more information.

12. Tanh: The Tanh function.

13. P_G: Probability of a generated sample.

Since the training data are normalized building power demands calculated using Equation 6, extra work needs to be done to convert generated data into building power demands. In this process, the maximum and minimum building power demands for each building type are needed. However, SGAN and InfoGAN do not provide information about building types for generated data. To solve this problem, we use the CNN classifier [41] with the Softmax function [45] from the SGAN model to calculate the probability that a generated sample belongs to each building type. Then, in our results, we select the building type with the highest probability as the building type for this sample. During the training process, the CNN classifier is able to successfully classify the real data with 95% accuracy within 150 epochs.

5.2. Results for Model Suitability

This subsection evaluates the model suitability for the five studied GAN types. We use the case for four building types shown in Table 2 to evaluate the model suitability. The training sample set for all models of the five studied GAN types contains 100 real building samples for each building type. All models for the five studied GAN types are trained for 2,000 epochs.

Then, the building power demand samples are generated using these trained GAN models. Figure 6 shows the number of generated samples. After the evaluation, SGAN and InfoGAN are unable to control the number of generated samples for each building type. The results show that SGAN generates power demand data for single-family detached house and multi-family low-rise apartment buildings, while it does not generate any samples for small office and stand-alone retail buildings.

Unlike the family of conditional GANs, the SGAN is **not** conditioned on a building type, making it difficult to generate an even number of samples per type. In the case of the GAN, there is one type of GAN per building type; so it is precisely known the generated sample's building type. While prior results from the SGAN demonstrates comprehensive coverage over the data distribution, our results, which were trained with prior methods, demonstrate that the SGAN developed a bias towards generating only two different building types. We also generated beyond the number of samples in the dataset (fourfold times the sample size) and were unable to generate samples from a different building type. The SGAN has no mechanism to specifically include building type. Even when training on a dataset with an even number of samples per type, the distribution is difficult to train on.

InfoGAN mainly generates power demand data for stand-alone retail buildings, with a small set of power demand data generated for small office buildings. However, it does not generate any data for the other two building types. We did not find any correlation between the latent code (which we set to 4 different codes) and the actual building type. Because of this, multiple category codes were actually associated with one building type, making it infeasible to generate certain samples for a specific building type. On the other hand, using building type as an input for the generator in the other three GAN types (Original GAN, cGAN, and ACGAN) allows them to generate the required number of building samples for each building type, which are listed in Table 2. Thus, SGAN and InfoGAN do not meet the first criterion in model suitability (subsection 3.1)that a GAN model must have the capability of generating the required number of building samples for each building type.

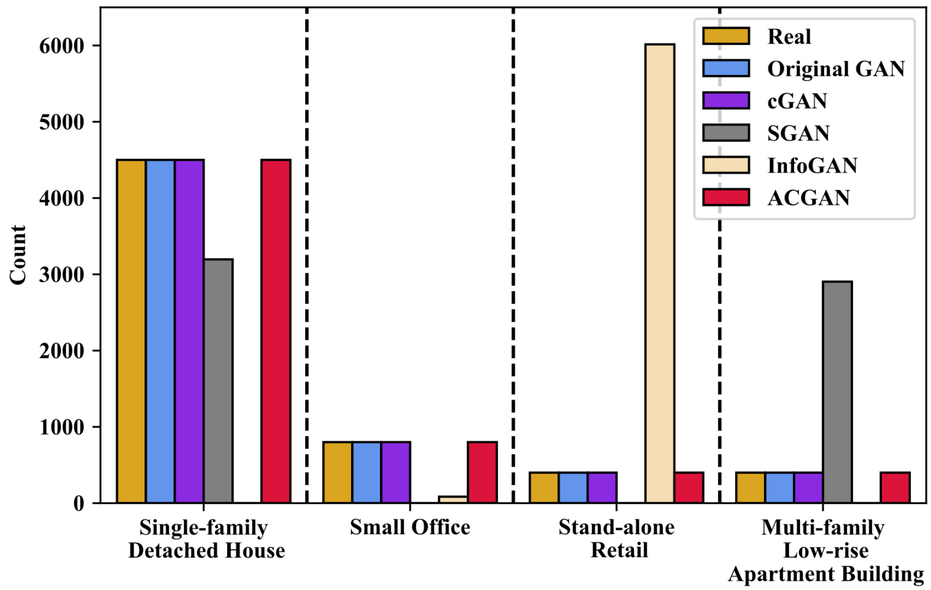

Figure 6: Number of generated samples for all studied GANs.

Figure 7 uses power demand in one day (August 1^{st}) to evaluate the feature capture for SGAN and InfoGAN. Since SGAN and InfoGAN only generate power demand data for two building types; this figure only displays the result for these building types. Figure 7(a) and (b) shows that the data generated by SGAN can capture the main features of real data for single-family detached house and multi-family low-rise apartment buildings since the generated samples are similar to real samples. This suggests that while bias in sample generation is apparent in the SGAN, it is still able to generate accurate samples. Figure 7(c) and (d) shows that the data generated by InfoGAN cannot capture the main features of real data for small office and stand-alone retail buildings since there are significant difference in power between real samples and generated samples. Even though the latent codes of the InfoGAN have shown to correspond with the actual class a sample belongs to, they did not correspond in our case. In the InfoGAN, there is no ability of facilitating the correlation between latent codes and actual building types. Thus, InfoGAN does not meet the second criterion in module suitability that the generated power demand data needs to capture the main features of real data.

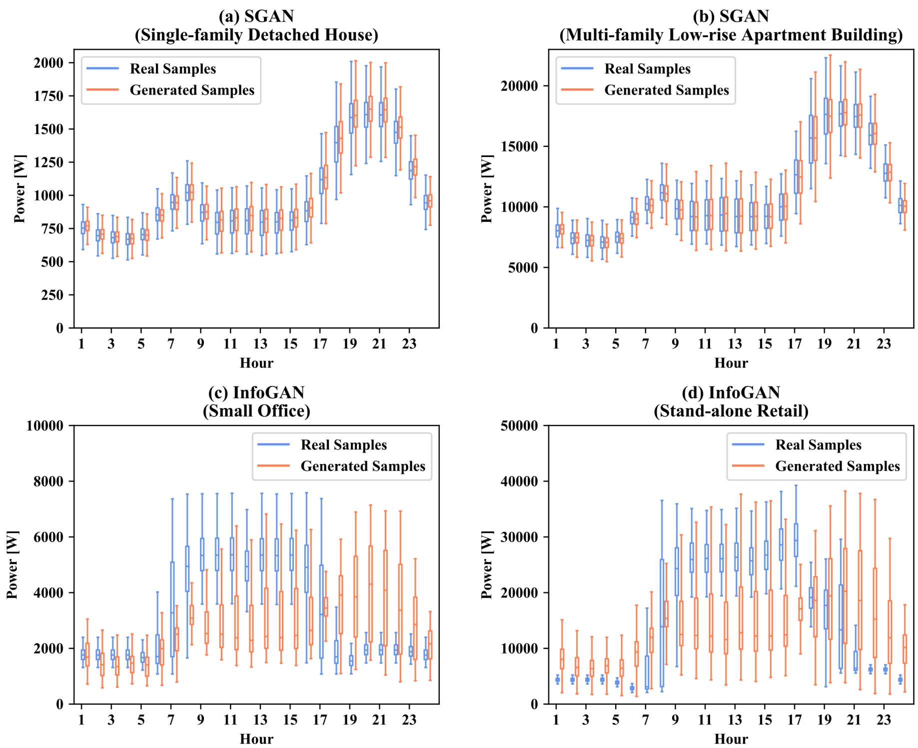

Figure 7: Feature capture for SGAN and InfoGAN by using the distribution of power data on August 1^st. *((a) power demand prediction for single-family detached house using SGAN; (b) power demand prediction for small office using InfoGAN; (c) power demand prediction for multi-family low-rise apartment buildings using SGAN; (d) power demand prediction for stand-alone retail using InfoGAN)*

Figure 8 uses power demand in one day (August 1^{st}) for single-family detached houses as an example to show the performance of power demand prediction generated by the models for Original GAN, cGAN and ACGAN. The distributions of power demands generated by the Original GAN models and cGAN model are both close to the real samples. The power demand data generated by the ACGAN model are able to capture the main features of the real data, but it tends to generate samples to reflect the median level of the distribution of the real data. We hypothesize that the ACGAN's **lack of auxiliary information provided to the discriminator** makes learning more challenging, whereas the CGAN has information provided to its discriminator.

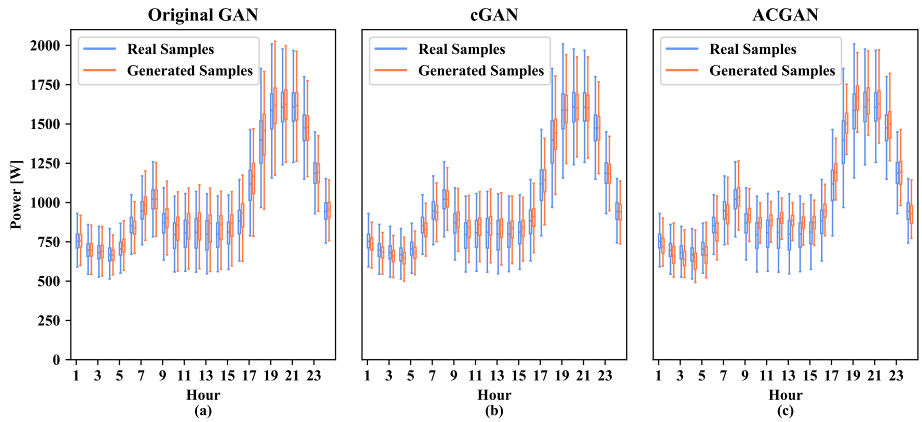

Figure 8: Feature capture for different types of GANs by using the distribution of power data for all generated single-family detached house samples on August 1^{st}.

The above results show that Original GAN, cGAN and ACGAN meet the module suitability criteria introduced in Subsection 3.1, but SGAN and InfoGAN fail in the test. Thus, Original GAN, cGAN and ACGAN will be studied in the following subsection for their model sensitivity to training size and number of building types.

5.3. Results for Model Sensitivity

The performance of GAN is highly impacted by training sizes (number of training samples). In this study, the Original GAN is used to select training sizes in the quantitative analysis. The numbers of training samples are randomly selected from 10 to 100 with an interval of 10. In addition, the epochs of Original GAN model training ranged from 50 to 2,000 with an interval of 50. For each training size and number of epochs, we train GAN models and generate building samples 30 times, which are the total number of runs.

Figure 9 shows $R^2_{k_1,k_2}$ for generated building samples by using the Original GAN models. The red and blue lines are the results with 30 and 100 training samples for each building type, respectively. The grey lines represent the remaining training sample sizes.

The upper left plot (k_1 = mean) shows that, when there are 30 and 100 training samples, $R^2_{k_1=mean,k_2}$ for k_2 (minimum, 25^{th} percentile, mean, median, 75^{th} percentile, and maximum) exceed 0.95 after 600 epochs, which

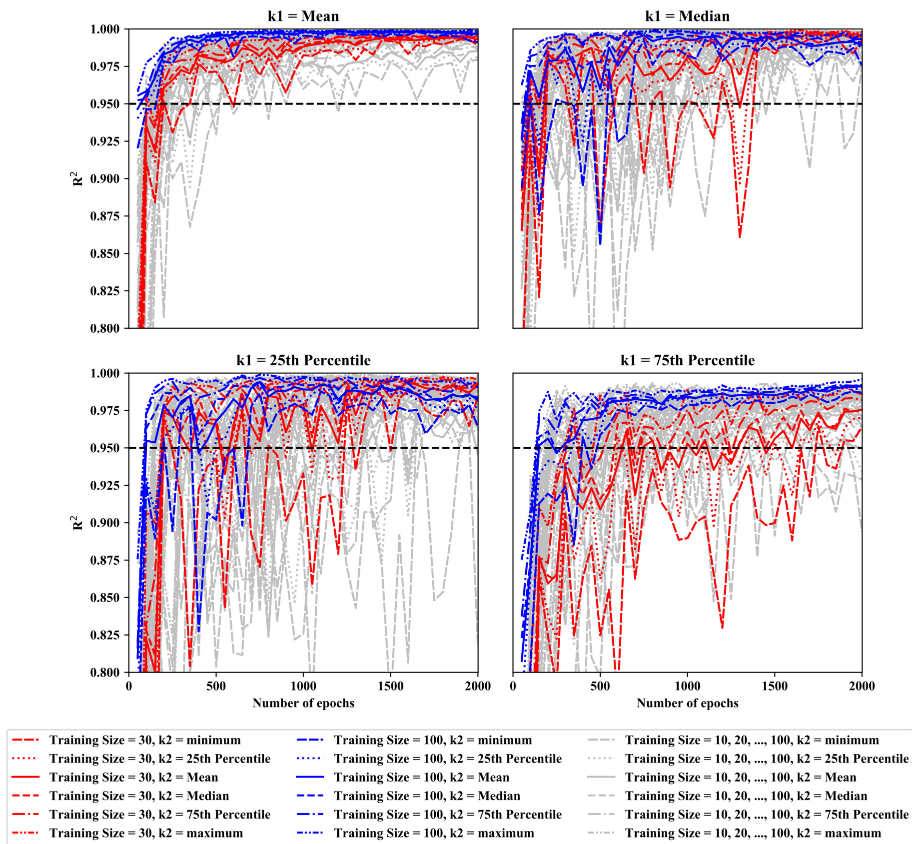

Figure 9: $R^2_{k_1,k_2}$ for generated building samples by using the Original GAN models.

means good prediction performance and a high number of credits for the evaluation indicator. The $R^2_{k_1=mean,k_2}$ for 100 training samples are higher than 30 training samples for all k_2 and all studied epochs. This is due to the extra noise in training from the lower amount of samples, which makes is more challenging for the generator to generalize across the whole dataset. The increased amount of samples reduces the noise and reflects the whole dataset better during training, leading to smoother updates of the model. The upper right plot (k_1 = median) shows that $R^2_{k_1=median,k_2}$ for 100 training samples are all higher than 0.95 after 1,400 epochs. However, $R^2_{k_1=median,k_2}$ for the set of 30 training samples has more fluctuation in different numbers of epochs than the one with 100 training samples. After 1,500 epochs, $R^2_{k_1=median,k_2}$ for 30 training samples become stable and all $R^2_{k_1=median,k_2}$ are higher than

0.95. The lower left plot ($k_1 = 25^{th}$ percentile) is similar to the upper right plot ($k_1 =$ median). $R^2_{k_1=25^{th}percentile,k_2}$ for 100 training samples is higher than 0.95 after 700 epochs, while $R^2_{k_1=25^{th}percentile,k_2}$ for 30 training samples fluctuates significantly. $R^2_{k_1=25^{th}percentile,k_2}$ for 30 training samples become stable and are all higher than 0.95 after 1,500 epochs. In the lower right plot ($k_1 = 75^{th}$ percentile), $R^2_{k_1=75^{th}percentile,k_2}$ for 100 training samples is higher than 0.95 after 500 epochs, while $R^2_{k_1=75^{th}percentile,k_2}$ for 30 training samples fluctuates. When the number of epochs increases, $R^2_{k_1=75^{th}percentile,k_2}$ for 30 training samples becomes higher, but can still be lower than 0.95 in some cases when k_2 is minimum or 25^{th} percentile. However, the noisier data of the 30 samples is still detrimental on the 75^{th} percentile.

After calculating $R^2_{k_1,k_2}$, the evaluation indicator (EI) is calculated. Figure 10 displays the evaluation indicators for all studied training samples and epochs by using the Original GAN models. The evaluation indicators for 30 and 100 training samples are highlighted, in which $R^2_{k_1,k_2}$ are shown in red and blue colors in Figure 10, respectively.

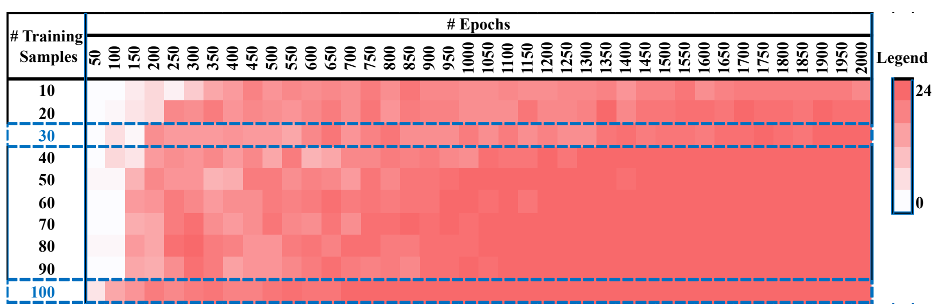

Figure 10: Evaluation indicator (EI) for generated building samples by using the Original GAN models.

In general, the evaluation indicators, introduced in Subsection 3.2.1, become higher when the number of training samples and epochs increase, which means the better performance. This follows standard machine learning training principles, which emphasizes on training large datasets for extended periods of time. For example, when the number of epochs increases, the evaluation indicator for 30 training samples becomes higher and reaches 24 (full credits) at 2,000 epochs. There is a better performance for 100 training samples than 30 training samples. After 700 epochs, all evaluation indicators for 100 training samples are all 24 (full credits).

In the performance evaluation, 30 and 100 training samples are selected. This paper uses a small set of building samples for one building type as an example to demonstrate the process.

Then, four building types (single-family detached house, small office, stand-alone retail, and multi-family low-rise apartment building) are selected to evaluate the performance for the three studied GANs (Original GAN, cGAN, and ACGAN) in predicting building power demand at a large scale. Figure 11 displays the EIs for generated building samples by using the GAN models for the three selected GANs.

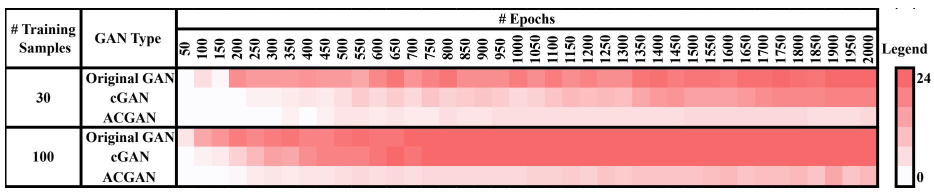

Figure 11: Evaluation indicator (EI) for generated building samples (four building types) by using the GAN models for the three selected GANs.

As discussed for Figure 11, the EIs for the Original GAN generally become higher when the number of training samples and epochs increase. The EIs for 100 training samples are all 24 after 700 epochs, which is better than 30 training samples. cGAN has worse performance than the Original GAN when the training size or number of epochs is small. However, when the number of epochs is higher than 750, the EIs for 100 training samples are all 24. ACGAN has the worst performance among the three types of GAN. The highest EI for ACGAN in those cases is only 14. As previously mentioned, the ACGAN discriminator does not receive auxiliary information about building type during training, which adds an extra challenge to the ACGAN and degrades performance.

This section also studies the impact of number of building types on the performance of different GANs, as shown in Figure 12. This figure displays the performance for four and six building types for Original GAN, cGAN, and ACGAN. The selection for different numbers of building types has been introduced in Table 2.

Since one Original GAN model is only used for one building type, the performance for the Original GAN is not affected by the number of building types. Figure 12 shows that the performance of the Original GAN in the studies for four and six building types is similar. It is because each Original

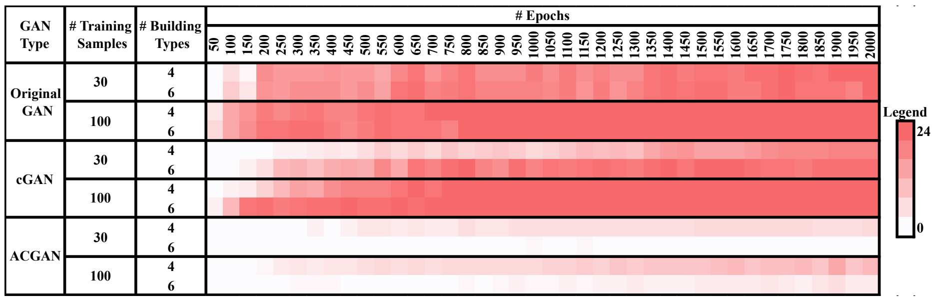

Figure 12: Impact on the number of building types (four and six building types).

GAN model is trained independently. However, the number of building types greatly affects the performance for cGAN and ACGAN. Interestingly, the performance of cGAN for six building types is better than four building types. This is due to the fact that the cGAN is able to easily learn how to accurate the other two building types, which improves the overall EI. The cGAN for six building types can more easily get close to 24 when the number of training samples or epochs is small. Different from cGAN, when the number of building types increases, the performance of ACGAN becomes worse. In the study for six building types, the highest EI for ACGAN is only five, again suggesting the degrading in performance of the ACGAN from a weaker discriminator.

To validate the findings for cGAN, the case for five building types is also used. Figure 13 displays the impact of the number of building types for cGAN, which considers the cases for four, five, and six building types. The selection for different numbers of building types has been introduced in Table 2.

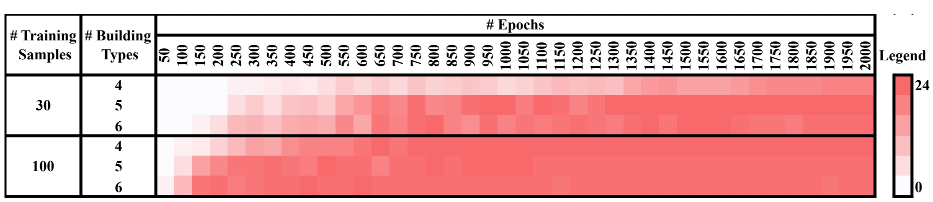

Figure 13: Impact on the number of building types for cGAN (four, five, and six building types).

When four building types are used and there are 30 training samples for each building type, cGAN can only obtain 20 credits for the EI after 2,000 epochs. In the cases for five and six building types, the EI for cGAN is 24 (full credits) after 1,300 epochs. When there are 100 training samples for each building type, the EI for cGAN is 24 for all cases (four, five, and six building types). However, when the number of building types increases, the EI for cGAN tends to be higher for small number of epochs (e.g., 200 epochs). In conclusion, increasing the number of building types improves the cGAN performance. This is because the data for different building types benefit from each other for predicting building power demand.

5.4. Discussion

This study finds that the SGAN and InfoGAN are not suitable for large-scale building power demand prediction. The Original GAN and cGAN are able to predict building power demand more accurately than ACGAN; the ACGAN's discriminator must learn extra information as compared to the GAN and CGAN. To use the Original GAN for the prediction, multiple GAN models need to be trained for different classes of buildings; to use the cGAN/ACGAN for the prediction, only one GAN model needs to be trained for all buildings, but the class for each training sample needs to be given. In this study, only 18 variables' uncertainties are considered and the shapes of building power demands for one building type are similar. Thus, the building type is a qualified classifier for the building samples. The actual cases contain more variability and randomness, and the shapes of building power demands for one building type vary case by case. In the actual study for building power demand prediction at a large scale, many building types and various power demand features are contained. To use GAN in the actual study, classification of buildings is needed and building type may not be the only variable to classify building samples. Many other variables, such as construction year, total floor area, and system type, can be candidates. Furthermore, statistical methods about classifying and clustering samples are also ways to classify building samples. In addition, the selection of representative building samples for GAN model training is also important, which should be taken into consideration when needed. Since cGAN has better performance when the number of building types increases, cGAN may be more qualified than the Original GAN when involving many building types with various power demand features.

6. Conclusion

This paper develops a novel strategy to evaluate the performance of different GANs for predicting building power demand at a large scale. Five promising GANs (Original GAN, cGAN, SGAN, InfoGAN, and ACGAN) are identified and evaluated using this strategy. Based on the result, Original GAN and cGAN are recommended, while SGAN, InfoGAN, and ACGAN are excluded from the candidate list for research about large-scale building power demand prediction. With a limited number of training samples, Original GAN provides better accuracy than cGAN. When the number of building types increases, the prediction accuracy increases for cGAN. In the future, the Original GAN and cGAN can be utilized into actual cases, which will be more complex than the studied cases in this paper. Their performance about building power prediction at a large scale will be studied in these actual cases.

References

[1] B. P. Center, Annual energy outlook 2020, Energy Information Administration, Washington, DC (2020).

[2] H. Hao, C. D. Corbin, K. Kalsi, R. G. Pratt, Transactive control of commercial buildings for demand response, IEEE Transactions on Power Systems 32 (2016) 774–783.

[3] N. W. Miller, M. Shao, S. Pajic, R. D'Aquila, Western wind and solar integration study phase 3–frequency response and transient stability, Technical Report, National Renewable Energy Lab.(NREL), Golden, CO (United States), 2014.

[4] F. Winkelmann, B. Birdsall, W. Buhl, K. Ellington, A. Erdem, J. Hirsch, S. Gates, DOE-2 supplement: version 2.1 E, Technical Report, Lawrence Berkeley Lab., CA (United States), 1993.

[5] D. B. Crawley, L. K. Lawrie, C. O. Pedersen, F. C. Winkelmann, Energy plus: energy simulation program, ASHRAE journal 42 (2000) 49–56.

[6] J. A. Ingraham, J. R. New, Virtual EPB, Technical Report, Building Technologies Office following the BTO Peer Review, 2018.

[7] Y. Chen, T. Hong, M. A. Piette, Automatic generation and simulation of urban building energy models based on city datasets for city-scale building retrofit analysis, Applied Energy 205 (2017) 323–335.

[8] B. Polly, C. Kutscher, D. Macumber, M. Schott, S. Pless, B. Livingood, O. Van Geet, From zero energy buildings to zero energy districts, Proceedings of the 2016 American Council for an Energy Efficient Economy Summer Study on Energy Efficiency in Buildings, Pacific Grove, CA, USA (2016) 21–26.

[9] J. Sanyal, J. New, R. E. Edwards, L. Parker, Calibrating building energy models using supercomputer trained machine learning agents, Concurrency and Computation: Practice and Experience 26 (2014) 2122–2133.

[10] Y. Ye, Y. Lou, M. Strong, S. Upadhyaya, W. Zuo, G. Wang, Development of new baseline models for us medium office buildings based on commercial buildings energy consumption survey data, Science and Technology for the Built Environment 26 (2020) 1321–1336.

[11] V. Cherkassky, S. R. Chowdhury, V. Landenberger, S. Tewari, P. Bursch, Prediction of electric power consumption for commercial buildings, in: The 2011 International Joint Conference on Neural Networks, IEEE, 2011, pp. 666–672.

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

[13] S. Seyedzadeh, F. P. Rahimian, P. Rastogi, I. Glesk, Tuning machine learning models for prediction of building energy loads, Sustainable Cities and Society 47 (2019) 101484.

[14] A. Nutkiewicz, Z. Yang, R. K. Jain, Data-driven urban energy simulation (due-s): A framework for integrating engineering simulation and machine learning methods in a multi-scale urban energy modeling workflow, Applied energy 225 (2018) 1176–1189.

[15] A. Nutkiewicz, Z. Yang, R. K. Jain, Data-driven urban energy simulation (due-s): Integrating machine learning into an urban building energy simulation workflow, Energy Procedia 142 (2017) 2114–2119.

[16] N. Abbasabadi, M. Ashayeri, R. Azari, B. Stephens, M. Heidarinejad, An integrated data-driven framework for urban energy use modeling (ueum), Applied energy 253 (2019) 113550.

[17] Z. Yang, J. Roth, R. K. Jain, Due-b: Data-driven urban energy benchmarking of buildings using recursive partitioning and stochastic frontier analysis, Energy and Buildings 163 (2018) 58–69.

[18] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative adversarial networks, arXiv preprint arXiv:1406.2661 (2014).

[19] J. F. Torres, A. M. Fernández, A. Troncoso, F. Martínez-Álvarez, Deep learning-based approach for time series forecasting with application to electricity load, in: International Work-Conference on the Interplay Between Natural and Artificial Computation, Springer, 2017, pp. 203–212.

[20] C. Tian, C. Li, G. Zhang, Y. Lv, Data driven parallel prediction of building energy consumption using generative adversarial nets, Energy and Buildings 186 (2019) 230–243.

[21] C. Tian, Y. Ye, Y. Lou, W. Zuo, G. Zhang, C. Li, Daily power demand prediction for buildings at a large scale using a hybrid of physics-based model and generative adversarial network, in: Building Simulation, Springer, 2022, pp. 1–17.

[22] J. Gui, Z. Sun, Y. Wen, D. Tao, J. Ye, A review on generative adversarial networks: Algorithms, theory, and applications, arXiv preprint arXiv:2001.06937 (2020).

[23] Z. Wang, T. Hong, Generating realistic building electrical load profiles through the generative adversarial network (gan), Energy and Buildings 224 (2020) 110299.

[24] X. Zhu, A. B. Goldberg, Introduction to semi-supervised learning, Synthesis lectures on artificial intelligence and machine learning 3 (2009) 1–130.

[25] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, X. Chen, Improved techniques for training gans, arXiv preprint arXiv:1606.03498 (2016).

[26] B. Li, F. Cheng, X. Zhang, C. Cui, W. Cai, A novel semi-supervised data-driven method for chiller fault diagnosis with unlabeled data, Applied Energy 285 (2021) 116459.

[27] B. Li, F. Cheng, H. Cai, X. Zhang, W. Cai, A semi-supervised approach to fault detection and diagnosis for building hvac systems based on the modified generative adversarial network, Energy and Buildings 246 (2021) 111044.

[28] D. Wu, K. Hur, Z. Xiao, A gan-enhanced ensemble model for energy consumption forecasting in large commercial buildings, IEEE Access (2021).

[29] X. Chen, Y. Duan, R. Houthooft, J. Schulman, I. Sutskever, P. Abbeel, Infogan: Interpretable representation learning by information maximizing generative adversarial nets, arXiv preprint arXiv:1606.03657 (2016).

[30] M. Mirza, S. Osindero, Conditional generative adversarial nets, arXiv preprint arXiv:1411.1784 (2014).

[31] C. Faulkner, D. Jankowski, W. Zuo, P. Epple, M. Sohn, A novel physics-informed algorithm for training ai models to predict indoor airflow, Proceedings of the 12th International Symposium on Heating, Ventilation and Air Conditioning (ISHVAC 2021) (2021).

[32] B. Tang, Y. Tu, Z. Zhang, Y. Lin, Digital signal modulation classification with data augmentation using generative adversarial nets in cognitive radio networks, IEEE Access 6 (2018) 15713–15722.

[33] A. Odena, C. Olah, J. Shlens, Conditional image synthesis with auxiliary classifier gans, in: International conference on machine learning, PMLR, 2017, pp. 2642–2651.

[34] J. Adler, S. Lunz, Banach wasserstein gan, Advances in Neural Information Processing Systems 31 (2018).

[35] X. J. Zhu, Semi-supervised learning literature survey (2005).

[36] N. M. M. Bendaoud, N. Farah, S. B. Ahmed, Comparing generative adversarial networks architectures for electricity demand forecasting, Energy and Buildings 247 (2021) 111152.

[37] U. DOE, Commercial prototype building models, https://www.energycodes.gov/prototype-building-models#Commercial, 2021.

[38] U. DOE, Residential prototype building models, https://www.energycodes.gov/prototype-building-models#Residential, 2021.

[39] EIA, Commercial buildings energy consumption survey (cbecs), US Department of Energy: Washington, DC, USA (2021).

[40] EIA, Residential energy consumption survey (recs), US Department of Energy: Washington, DC, USA (2021).

[41] N. Kalchbrenner, E. Grefenstette, P. Blunsom, A convolutional neural network for modelling sentences, arXiv preprint arXiv:1404.2188 (2014).

[42] CS231n, Convolutional neural networks for visual recognition, https://cs231n.github.io/convolutional-networks, 2021.

[43] J. Brownlee, How to develop a gan for generating mnist handwritten digits, Internet: https://machinelearningmastery. com/how-to-develop-a-generative-adversarial-network-for-an-mnist-hand written-digits-from-scratch-in-keras/June 28 (2019).

[44] D. into Deep-Learning, Transposed convolution, https://d2l.ai/chapter_computer-vision/transposed-conv.html, 2021.

[45] I. Goodfellow, Y. Bengio, A. Courville, Softmax units for multinoulli output distributions. deep learning, 2018.