Journal of Combinatorial Mathematics
and Combinatorial Computing
www.combinatorialpress.com/jcmcc

Combinatorial Press

Designing Novel Biofuels Using Generative Adversarial Networks

Xiqi Yang[1,✉]

[1] *Faculty of Maths & Physical Sciences, University College London, London, United Kingdom*

ABSTRACT

New biofuels, as a sustainable energy alternative to traditional fossil fuels, are attracting global attention. With the increasing awareness of environmental protection and the continuous growth of energy demand, biofuels offer the possibility of reducing greenhouse gas emissions and decreasing dependence on fossil fuels. In this paper, by introducing the Wasserstein distance, which is used to describe the objective function of the GAN model, the self-attention mechanism is applied to improve the discriminator structure of the traditional WGAN-GP to achieve more efficient generation of high-quality data samples. The WGAN-GP model is used to design a new biofuel combustion scenario, and based on the combustion data, the new biofuel is prepared in the scenario. The final data generation results of the model are evaluated based on relevant evaluation indexes. It can be seen that the trend of the generated data set is consistent with the trend of the actual output value of the power station, and the interval range formed by the generated 50 sets of data can include the real data in a more complete way, with a high data coverage, and the error between the generated value and the real value is in the range of ±250-±300. The new biofuel output scenarios generated by the WGAN-GP model were utilized for EMF synthesis experiments. PTFE@ACMS-SO_3H samples showed strong absorption peaks at 759 cm^{-1} and 54 cm^{-1}, indicating that the acidic groups-SO_3H were successfully loaded on the surface of the material and the preparation of the novel biofuel was successful.

Keywords: Wasserstein distance, self-attention mechanism, WGAN-GP model, novel biofuel

1. Introduction

Novel biofuels, as a sustainable energy alternative to traditional fossil fuels, are attracting global attention. With increased awareness of environmental protection and continued growth in energy demand, biofuels offer the potential to reduce greenhouse gas emissions and decrease dependence on fossil fuels [1-3]. Technological advances have driven the shift from the first generation of biofuels relying on food crops to advanced forms utilizing non-edible biomass, algae, and even gene editing and

✉ Corresponding author.
E-mail address: yxq15110374733@163.com (X. Yang).
Received 12 February 2024; Revised 25 May 2024; Accepted 30 September 2024; Published Online 16 April 2025.
DOI: 10.61091/jcmcc127b-020

nanotechnology, which not only indicate technological innovations, but also reflect in-depth consideration of environmental sustainability [4-6].

The categorization of new biofuels demonstrates the shift from traditional to more sustainable energy sources. In this field, biofuels are categorized into four generations according to their production sequence and feedstock type, with each generation representing an evolution in technology and a deepening of environmental considerations [7-9]. First, there is the first generation of biofuels, produced mainly from food crops such as corn, sugarcane and soybeans, which are converted into ethanol and biodiesel through a fermentation or refining process. This generation of biofuels has a simple production process, but its sustainability has been questioned due to competition with food supplies for land and resources [10-12]. In response to this challenge, the second generation of biofuels has emerged, utilizing mainly non-food plants or waste biomass, such as straw, bark and waste plants. The use of these feedstocks reduces the impact on food supply and biodiversity and increases the environmental friendliness of biofuels [13-15]. The third generation of biofuels, which follows closely behind, uses specially cultivated algae as feedstock. The high lipid content of algae makes it an ideal feedstock for biodiesel production. More importantly, algae can be grown on non-arable land, further alleviating the need for agricultural land. At the forefront is the fourth generation of biofuels, which combines gene editing and nanotechnology [16-18]. This generation of biofuels not only extracts energy from biomass, but also effectively captures and stores atmospheric carbon dioxide to combat climate change. With the evolution from the first to the fourth generation, biofuels have made significant progress in reducing dependence on food crops, lowering the impact on ecosystems, and increasing energy efficiency [19-21]. This evolution reflects not only technological developments but also the increasing emphasis on environmental sustainability and energy security [22].

Gasoline and diesel are widely used automobile engine fuels, and they are both products made from petroleum. However, in recent years, petroleum resources have become increasingly tight and are no longer able to meet the global demand for energy consumption, which has led to a surge in petroleum prices [23-24]. The Chinese government advocates the establishment of a conservation-oriented harmonious society to cope with the impact of energy constraints on China's economic growth, and at the same time, people are paying more and more attention to the quality of life and environmental protection, and therefore require new alternative energy sources to meet the conditions of renewable and non-hazardous to human health and environmental protection, which is precisely why the research and development of renewable and environmentally friendly biofuels have been rapidly developed [25-27].

Literature [28] describes the recent advances in microbial fuel technology and potential future application areas, and points out that the focus of research on microbial fuels should be on biofuel-producing species, optimization and improvement of cultivation conditions, genetic engineering, and so on. Literature [29] analyzed the issues related to the feasibility and commercialization of biofuels in India and explored whether the development of biofuel technology can meet the energy needs of India. Literature [30] reviewed the application of nanotechnology in biofuel research and pointed out that nanotechnology is extremely valuable for the commercial development of biofuels and also helps in environmental protection. Literature [31], through a systematic review of biofuel research papers, revealed that lignocellulosic biomass can produce bioethanol technology pathways, and uncovered the potential barriers, and targeted recommendations to improve the efficiency of bioethanol fuel production. Literature [32] describes the positive significance of biofuels for environmental protection and elucidates the potential of microorganisms such as Escherichia coli and Clostridium acetobutylicum for the production of biobutanol fuels from waste carbohydrates and the feasibility of the corresponding technological strategies. Literature [33] evaluated the RenovaBio project in terms of techno-economic and environmental feasibility and showed that the rational integration of microalgae fuel production with traditional sugarcane mills is economically feasible and can

significantly reduce greenhouse gas emissions. Literature [34] examined the financial ownership (CBIOs) acquired by Brazilian biofuel producers based on the ratio of biofuel production and efficiency, and concluded that the CBIOs have significant appreciation and investment value, with an outlook for a significant increase in sustainable biofuel production in Brazil in the future. Literature [35] examined biodiesel combustion performance and emissions and found that biodiesel has the largest soot gas emissions and the least NOX emissions, while an increase in unsaturated fatty acid content promotes an increase in soot output value and an increase in the hydrocarbon ratio significantly reduces HC and CO emissions.

In this paper, the generative adversarial network method is utilized to capture the data distribution, set the loss functions of the generator and the discriminator, and improve the discriminator structure of the traditional WGAN-GP by self-attentive modeling. Meanwhile, based on the traditional GAN, Wasserstein distance is introduced to describe the objective function to improve the stability of training and complete the transformation of the objective function. The WGAN-GP model is used as the basis to migrate the data of novel biofuel combustion scenarios. After game training, the random noise convolution value is calculated to generate the corresponding new biofuel combustion capacity data. Evaluation indexes such as RMSE and MAE are proposed to be calculated, and the effect of new biofuel combustion scenario data migration is assessed by evaluation indexes. Using the improved WGAN-GP model, the preset biofuel combustion scenarios and designing the novel biofuel based on the combustion data of the biofuel, the novel biofuel is prepared by catalytic synthesis EMF technology, and experiments are used to simulate the catalyst use process and describe the catalyst characterization structure.

2. Generation of novel biofuel combustion scenarios based on generative adversarial networks

2.1. Improved WGAN model

2.1.1. Generating Adversarial Networks. GAN is an unsupervised neural network that can implicitly learn the intrinsic probability distribution of the data by training the real dataset to continuously fit the model, and its basic structure is shown in Fig. 1.GAN is widely used to solve the problem of generating power data samples such as new energy scenario generation, unit fault diagnosis, and so on, because of its ability to generate a large amount of similar data when the quality of the original data is not high and to improve the quality of the generated data.

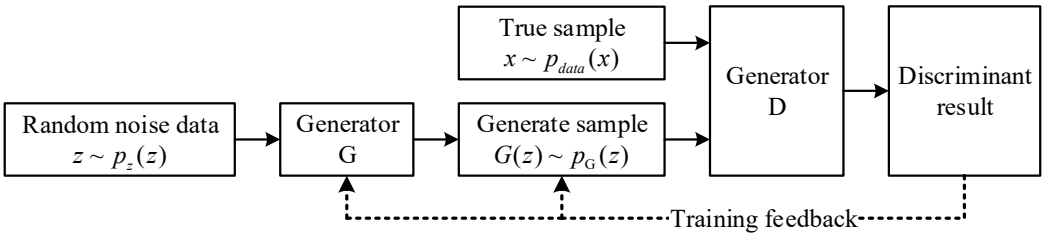


Fig. 1. Basic structure diagram of GAN model

In Fig. 1, generator G is used to capture the data distribution, discriminator D is used to determine the probability that the generated samples are from real data, and the two models are trained simultaneously. The training objective of generator G is to use the generated data to deceive discriminator D and maximize the probability of discrimination error of discriminator D. The loss functions of generator and discriminator are:

$$Loss_G = -\underset{z\sim p_l(z)}{E}\{D[G(z)]\} \tag{1}$$

$$Loss_0 = -\left\{ \underset{z\sim p_i(z)}{E}[\lg D(x)] + \underset{z\sim p_i(z)}{E}\{\lg\{1-D[G(z)]\}\} \right\} \tag{2}$$

Where: $E(\cdot)$ is the expectation function, $G(\cdot), D(\cdot)$ is the generator and discriminator functions, respectively, and $p_z(z), p_{data}(x)$ is the distribution of input noise z and sample input x, respectively.

The training process of the generator and discriminator is a max-min game with the objective function:

$$\min_G \max_D V(D,G) = \underset{z\sim p_{data}(z)}{E}\left[\lg D(x)\right] + \underset{z\sim p_z(z)}{E}\left\{\lg\left\{1-D\left[G(z)\right]\right\}\right\} \tag{3}$$

$V(D,G)$ is a dichotomous cross-entropy function. The ultimate goal of Eq. (3) is to minimize the JS distance between the generated sample probability distributions $p_G(z)$ and $p_{data}(x)$.

The only solution in the space of arbitrary functions G and D such that G reproduces the distribution of the training data is $D=0.5$. There is no need for any Markov chain or unfolded approximate inference network throughout the training process. And when the generator G and the discriminator D consist of multilayer perceptrons, the whole system can be trained by backpropagation.

2.1.2. Improved SA/WGAN models:

1) WGAN-GP. The original GAN uses the JS distance to describe the discriminator's objective function in Eq. (3), but at the beginning of training, the probability distributions of the generated samples and the input samples may not be intersected, resulting in a constant JS distance of lg2 and the problem of vanishing gradient. The problem of vanishing or exploding gradients that may occur during training is solved by introducing the Wasserstein distance to replace the JS distance in the original GAN and replacing the weight trimming with a gradient penalty to make the training more stable by limiting the gradient not to exceed the Lipschitz constant of K_0 [36]. The objective function of the optimized WGAN-GP model is [37]:

$$\begin{aligned}\min_G \max_D V(D,G) = \underset{z\sim p_{data}(z)}{E}[D(x)] - \underset{z\sim p_z(z)}{E}[D(G(z)] \\ -\lambda E\cdot[\square\nabla D(x') - K_0\square]^2\end{aligned} \tag{4}$$

Among them:

$$x' = \varepsilon x + (1-\varepsilon)G(z) \tag{5}$$

Where: ε is the random number, λ is the gradient penalty weight coefficient, which is generally taken as 10.

However, due to the slow convergence speed of the improved WGAN-GP, it is suitable for scenarios that do not require high training speed.

2) Self-attention model. Although the recurrent neural network based on convolutional kernel can theoretically establish longer-distance dependencies, it is practically difficult to establish long-distance dependencies due to the problem of information transfer capacity and the possible phenomenon of gradient disappearance.

Currently, long-distance dependencies are mainly established by increasing the number of layers of the network and using fully connected layers, but both methods have drawbacks. Increasing the number of layers of the network will lead to slower network learning, and the fully connected layer cannot deal with longer input sequences, so both methods are not good for directly establishing long-distance dependencies.

The self-attention mechanism is free from the traditional recurrent neural network (RNN), convolutional neural network (CNN), and long short-term memory neural network (LSTM), and can better capture the global pairwise interactions. Let the input sequence $X=[x_1,\cdots,x_N]\in R_{D_x\times N}$, the

output sequence after automatic assignment be $H=[h_1,\cdots,h_N]\in R_{D_v\times N}$, and Q,K,V be three different spaces that each input is mapped to: query space, key space, and value space, respectively, as follows [38]:

$$\begin{cases} Q=W_qX\in R_{D_k\times N} & W_q\in R_{D_k\times D_x} \\ K=W_kX\in R_{D_k\times N} & W_k\in R_{D_k\times D_x} \\ V=W_vX\in R_{D_v\times N} & W_v\in R_{D_v\times D_x} \end{cases} \tag{6}$$

Where: W_q,W_k,W_v is the parameter matrix of the outer linear mapping, D_x is the number of rows of the input matrix, N is the number of columns of the input matrix, D_k is the number of rows that map the input to the query space and key space, and D_v is the number of rows that map the input to the value space.

Mapping each input into 3 different spaces: query space, key space, and value space yields 3 different vectors: query vector $q_i\in R_{D_k}$, key vector $k_i\in R_{D_k}$, and value vector $v_i\in R_{D_v}$. $Q=[q_1,\cdots,q_i,\cdots,q_N]$ is the matrix composed of query vectors, $K=[k_1,\cdots,k_i,\cdots,k_N]$ is the matrix composed of key vectors, and $V=[v_1,\cdots,v_i,\cdots,v_N]$ is the matrix composed of value vectors.

The key-value pair format is used to represent the n th query vector $q_n\in Q$, where the keys are used to compute the attention distribution α_i and the values are used to compute the aggregated information. Use $(K,V)=[(k_1,v_1),\cdots,(k_N,v_N)]$ to represent the N set of input information of the query vector and the output attention function h_n is:

$$\begin{aligned} h_n &= att\left[(K,V),q_n\right] \\ &= \sum_{j=1}^{N}\alpha_{nj}v_j \\ &= \sum_{j=1}^{N} soft\max\left[s(k_j,q_n)\right]v_j \end{aligned} \tag{7}$$

Where: $att(\cdot)$ is the attention function, s(-) is the attention scoring function in the form of scaled dot product model, n,j is the position of the input and output vector sequences, respectively, $n\in[1,N],j\in[1,N]$, α_{nj} is the weight of the n th output attention to the j th input, and $softmax(\cdot)$ is the column-wise normalization function.

If the scaled dot product is used to represent the attention scoring function, the scoring function can be obtained more efficiently and its output vector sequence can be simplified as:

$$H=Vsoftmax(\frac{K^{\mathrm{T}}Q}{\sqrt{D_k}}) \tag{8}$$

$softmax(\cdot)$ such that each column of its output vector H has a value between 0 and 1.

The computation of the attention function in Eq. (6) shows that all input vectors are involved in the operation, capturing the global pairwise interaction relations. The schematic diagram of the self-attention model capturing the global pairwise interaction relation is shown in Fig. 2. The self-attention model can be used as a layer in a neural network, either instead of the traditional RNNCNN or alternatively with RNN/CNN.

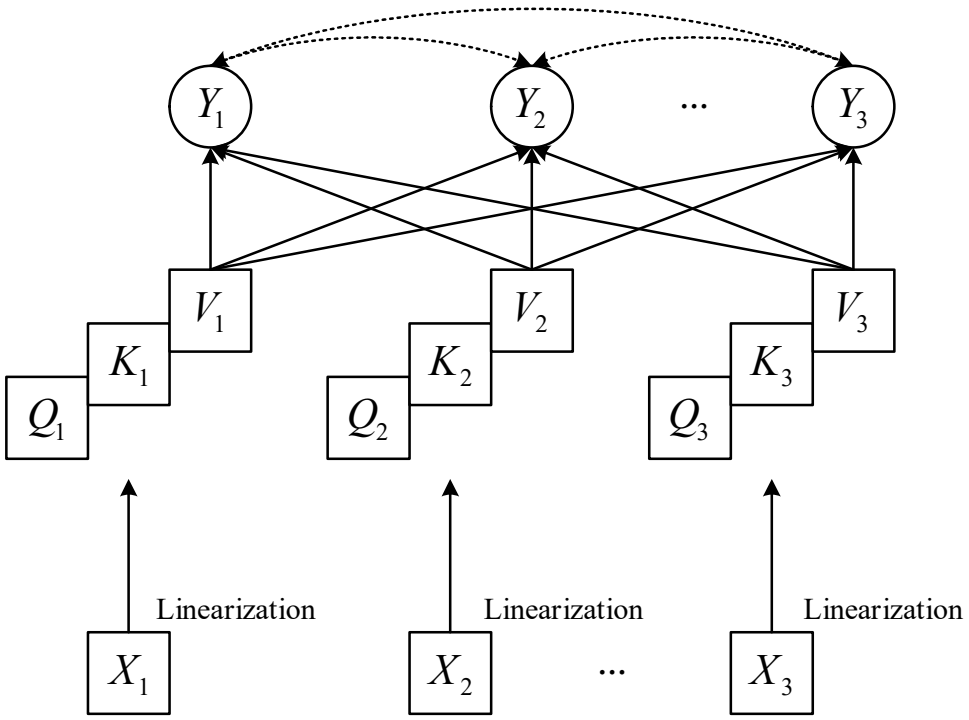


Fig. 2. Schematic diagram of the global pair-to-pair interaction captured by the self-attention

2.1.3. Improved WGAN-GP based on self-attention modeling. The self-attention model is used to improve the discriminator structure of the traditional WGAN-GP, which improves the convergence speed of the network and can generate high-quality data samples more efficiently and stably, and to design the network structure applicable to the generation of new biofuel combustion scenarios according to the data characteristics of the new biofuel combustion scenarios, and ultimately the improved WGAN-GP model based on the self-attention model used in this paper is proposed.

2.2. Data generation for novel biofuel combustion scenarios

2.2.1. Conditional Deep Convolutional Generative Adversarial Modeling. Define the historical novel biofuel combustion scenario data as real data, and for the generator network, define a set of random noise data z as the input of the generator, denote the probability distribution of z by $p_z(z)$, and denote the probability distribution of the historical novel biofuel combustion scenario data x by $p_{data}(x)$ at the same time. The output of the generator is the generated data sample $G(z)$ with probability distribution $p_G(z)$. The training objective of the generator is to make the probability distribution $p_G(z)$ of the generated scenario data samples as identical as possible to the real distribution $p_{data}(x)$.

For a discriminator network, the input is either historical scene data x or generator-generated data $G(z)$, and the output is a scalar $D(G(z))$ indicating the probability that the input data obeys the historical data distribution $p_{data}(x)$. The training objective of the discriminator is to discriminate whether the input data is historical or generated data as correctly as possible.

Based on the training objectives of the generator and the discriminator, the loss functions L_G and L_D of the generator and the discriminator, respectively, are constructed as follows:

$$L_G = -E_{z\sim p_z(z)} g\left[\log(1-D(G(z)))g\right] \tag{9}$$

$$L_D = -E_{x\sim p_{data}(x)}\left[\log D(x)\right] + E_{z\sim p_z(z)}\left[\log(1-D(G(z)))\right] \tag{10}$$

The optimization objective of the generator is to minimize Eq. (9) and the objective of the discriminator is to maximize Eq. (10). Combining Eq. (9) and Eq. (10) yields the objective function for the GAN training process, as shown in Eq. (11):

$$\min_G \max_D V(D,G) = E_{x\sim p_{dau}(x)}\left[\log D(x)\right] + E_{z\sim p_z(z)}\left[\log(1-D(G(z)))\right] \tag{11}$$

Since the original GAN is using JS distance to describe the objective function of Eq. (11). When the model just started training, the probability that the probability distribution of the generated data samples and the probability distribution of the historical data samples have no intersection at all is very high. And in the statistical theory analysis, when the two probability distributions have no intersection at all, the JS distance is constant $\log 2$. At this time, the objective function will not be able to accurately measure the distance between the sample distributions, which affects the accuracy of the data generation of the new bio-combustion scenarios.

Therefore, on the basis of traditional GAN, Wasserstein distance is introduced to describe the objective function in Eq. (11), which helps to solve the problem of gradient disappearance in training and improve the stability of training.Wasserstein distance is defined as follows:

$$W(p_{data}, p_G) = \inf_{\gamma\sim\Omega(p_{data},p_G)} E_{(u,v)\sim\gamma}\left[\| u-v \|\right] \tag{12}$$

where $\Omega(p_{data}, p_G)$ is the set of joint probability distributions γ with p_{data} and p_G as marginal distributions, and $W(p_{data}, p_G)$ is the lower bound of the $\gamma(u,v)$ expectation, which represents the distance that needs to be shifted from i to v to fit the generating distribution p_G to the true distribution p_{data}, where u and v denote the randomly sampled historical samples and generating samples from the joint distribution γ, respectively. Since it is difficult to obtain the Wasserstein distance by direct computation, its Kantorovich-Rubinsteim dyadic form is usually used to describe the distance between the generating sample and the true sample:

$$W(p_{data}, p_G) = \frac{1}{K}\sup_{\|f_D\|L\le K} E_{x\sim p_{data}}[D(x)] - E_{x\sim p_G}[D(x)] \tag{13}$$

Eq. $\| f_D \| L \le K$ shows that the discriminant function $D(x)$ satisfies K-Lipschitz continuity, i.e., the absolute value of the function gradient is capped at K. To ensure that the gradient does not exceed the limit K, a gradient penalty function for discriminant function $D(x)$ in the domain of definition is introduced to Eq. (11), so that the discriminant function $D(x)$ satisfies the K-Lipschitz continuity approximately to accurately describe the Wasserstein distance. At this point, the objective function of the GAN is transformed into:

$$\begin{aligned} &\min_G \max_D V(D,G) = E_{x\sim p_{dan}(x)}[D(x)] \\ &-E_{z\sim p_z(z)}[D(G(z))] - \lambda E[\| \nabla D(\sim) \| - K]^2 \end{aligned} \tag{14}$$

With the above improvements to the GAN objective function and loss function, the discriminator can not only distinguish between generated data samples and historical data samples, but also fit the Wasserstein distance between different sample sets and analyze the differences between different distributions. This difference can provide accurate training pointers for the generator, so that the probability distribution of the generated data is as much as possible the same as the probability distribution of the historical data, rather than simply generating data similar to the historical samples.

In this chapter, the above WGAN-GP model will be used as the basis for the research on data migration methods for novel biofuel combustion scenarios.

2.2.2. Data generation for new biofuel combustion scenarios. The new biofuel combustion scenario data as condition y is spliced with random noise z and input into the generator, enabling the convolutional layer of WGAN-GP to effectively analyze the local information features of the input data.

The convolutional layer, as the basic structure of WGAN, usually contains multiple convolutional kernels, which enables WGAN-GP to extract complex data features more comprehensively and quickly. The results generated by condition y, random noise z after convolutional kernel are shown in equations (15) and (16):

$$m_{yp} = f(w' \cdot y_{p:p+g-1} + b') \tag{15}$$

$$m_{zp} = f(w' \cdot z_{p:p+g-1} + b') \tag{16}$$

Where m_{yp}, m_{zp} denotes the condition y, random noise z the p th feature calculated by convolution respectively, f is the activation function set by the convolutional layer, w and b represent the weight and bias of each convolutional kernel respectively, and the window size of the convolutional kernel is g, and y_p, z_p is the p th input of the input feature $y, z, x, G(z)$.

Each convolutional layer is arranged with a number of convolutional kernels, and all the convolutional kernel features within the convolutional layer are spliced together, that is, the generator's convolutional layer output feature M_G, as shown in equation (17):

$$M_G = f(m_{yp:l}, L, m_{yp:n-g+l}, m_{zp:l}, L, m_{zp:n-g+l}) \tag{17}$$

From Eq. (17), after the convolution operation, the convolution layer output features are able to characterize the mapping relationship between the novel biofuel combustion data y and the random noise z. The convolution layer features extracted from the source power station data y and random noise z are continuously refined through multi-layer convolution calculation, and the generated sample data $G(z)$ is finally output.

The new biofuel combustion scenario data is used as the condition y, which is spliced with the generated sample $G(z)$ and the historical sample x, respectively, and inputted into the discriminator. The results generated by history sample x and generation sample $G(z)$ after convolution kernel are shown in equations (18) and (19):

$$m_p = f(w' \cdot z_{p:p+g-1} + b') \tag{18}$$

$$m_{Gp} = f(w' \cdot G(z)_{p:p+g-1} + b') \tag{19}$$

where m_{xp}, m_{Gp} denotes the p th feature calculated by convolution for history sample x and generation sample $G(z)$, respectively, and $x_p, G(z)_p$ is the p th input for input feature $x, G(z)$.

After multiple convolution kernels are computed, the output features M_{D1}, M_{D2} of the generated data $G(z)$ and the history sample x through the discriminator convolution layer are shown in equations (20) and (21), respectively:

$$M_{D1} = f(m_{\gamma p:1}, L, m_{\gamma p:n-g+1}, m_{Gp:1}, L, m_{Gp:n-g+1}) \tag{20}$$

$$M_{D2} = f(m_{yp:1}, L, m_{yp:n-g+1}, m_{xp:1}, L, m_{xp:n-g+1}) \tag{21}$$

2.2.3. Evaluation indicators. In order to verify the migration effect of the novel biofuel combustion scenario data, some statistical metrics were selected in this chapter to compare the real novel biofuel combustion scenario data with the migration-generated scenario data. The statistical indicators include mean, maximum, median, and standard deviation, and the accurate statistical indicators of the real scenario data are used as benchmarks to evaluate the migration-generated biofuel combustion scenario data. The above metrics reflect the distribution of the scenario data and the magnitude of the values, and the closer the metrics of the migrated generated biofuel combustion scenario data are to the real scenario data, the more accurate the migration proves to be.

In addition to this, the effect of migration of the scenario data was further evaluated by utilizing the root mean square error (RMSE), the mean absolute error (MAE), and the coefficient of determination R^2. The RMSE, MAE, and R^2 are defined as:

$$\mathrm{RMSE} = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(\hat{x}_t - x_t)^2} \tag{22}$$

$$\mathrm{MAE} = \frac{1}{T}\sum_{t=1}^{r}|\hat{x}_t - x_t| \tag{23}$$

$$R^2 = \frac{\sum_{t=1}^{T}(\hat{x}_t - \overline{x})^2}{\sum_{t=1}^{T}(x_t - \overline{x})^2} \tag{24}$$

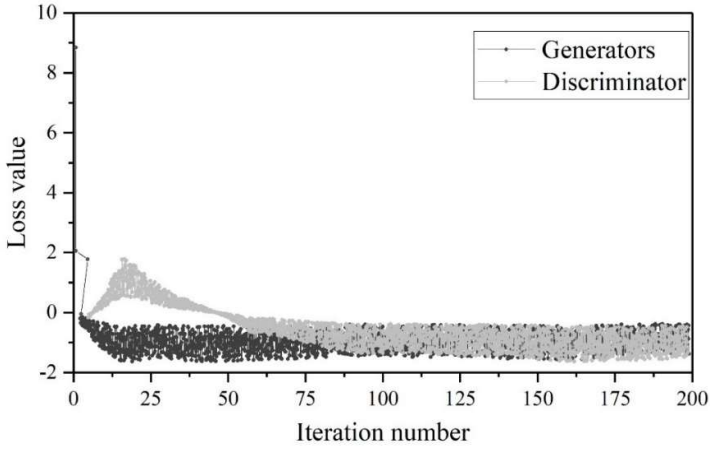
where T represents the total number of individual migrated scenario data, $\hat{x}_i, x_i$ and $\overline{x}_i$ represent the average of migrated biofuel combustion scenario numerical sampling points, real scenario numerical sampling points, and real scenario data, respectively [39].The RMSE and MAE are used to evaluate the error between migrated scenario data and real scenario data, and the lower the error is, the more accurate the migrated data are. The closer the coefficient of determination R^2 is to 1, the more correctly the migrated scene data represents the real scene data.

2.3. Example analysis

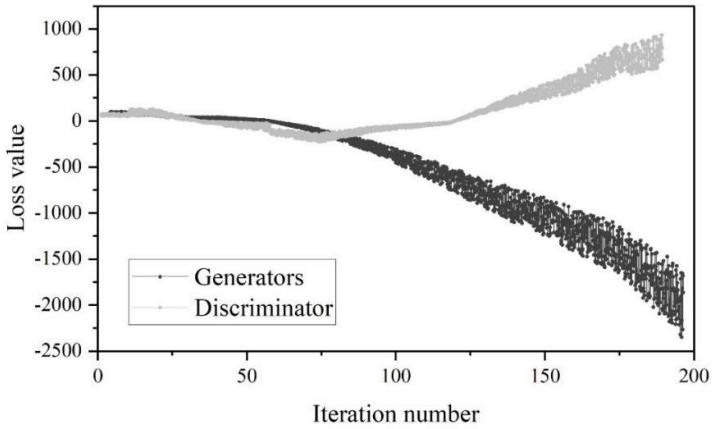
2.3.1. Loss function. The Wasserstein WGAN-GP model is applied to simulate a novel biofuel combustion scenario, and model training is performed below. The dataset used in this paper comes from the new biofuel combustion scenario data counted by transmission system operator EirGrid in Northern Ireland in 2021~2022. In addition, due to the adopted real dataset, there are inevitably missing data and erroneous data, and the samples with excessive deviation are removed by sample cleaning method, and the interpolation method is used to fill in the missing samples. As the selected samples have a certain degree of relevance, the batch training method is used for training.

After training, the network structure and parameters of the generator and discriminator are saved, and the conditional data y in the test set is spliced with Gaussian white noise z and input into the generator to obtain the new energy output value of the power station under the condition. The model loss function is shown in Fig. 3, Fig. 3(a) represents the generator and discriminator training loss function of the model in this paper, the loss function difference between the generator and discriminator is large at the beginning of the iteration, and the loss function value of the generator shows a trend of elevation, which is because at this time, there are large differences between the generator-generated data and the real data, and the Wasserstein distance between them is increasing [40]. As the iteration process continues, it can be found that the loss function of the discriminator decreases rapidly, and the loss function of the generator gradually decreases, which indicates that the data generated by the generator in the process of continuous iteration in the process of the generator in the process of constantly approximating the real data, in the iteration number of 100 times to start, the loss function of the generator and the discriminator gradually tends to stabilize, and in the axes near the value of 0 to reach the Nash equilibrium. Figure 3(b) represents the values of the generator loss function and discriminator loss function of the traditional model when the penalty term is not considered. It can be seen that as the number of iterations increases, the value of the generator's loss function increases, with the loss value going from 0 to near -2500, and the value of the discriminator's loss function decreases, going from 0 to near 500. Both loss functions diverge, which indicates that the constructed model is difficult to reach the Nash equilibrium near a constant value. Therefore, the

Wasserstein WGAN-GP loss function model with penalty term proposed in this paper has better stability as well as convergence compared to the traditional loss function model without penalty.



(a)This model

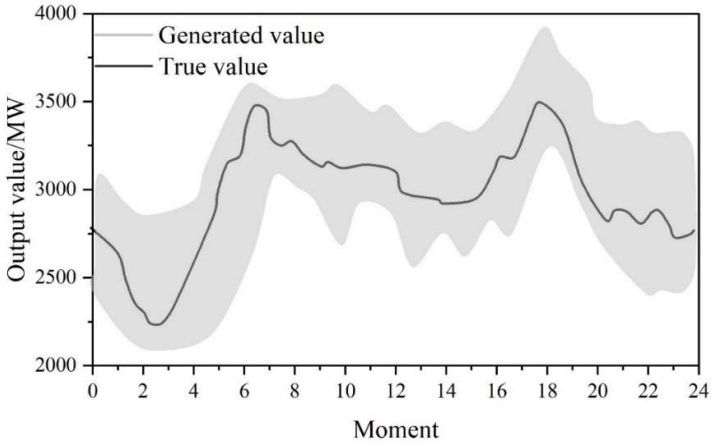


(b)Traditional model

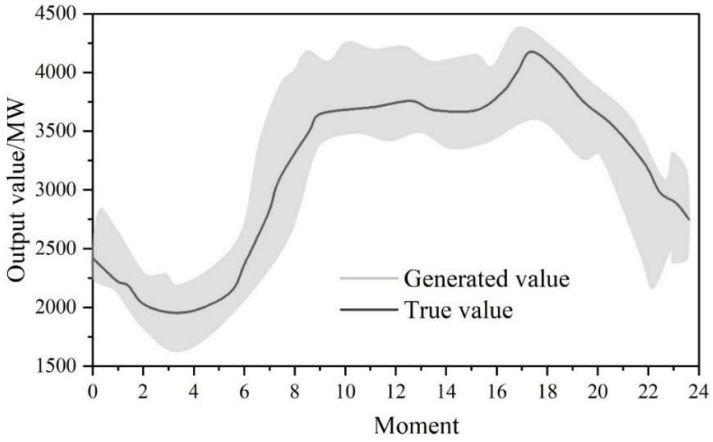
Fig. 3. Model loss function

2.3.2. Data generation results. After the training is completed, the amount of energy gained in the test set and the regional electrical energy demand are input into the generator as condition values spliced with noise z. The generator generates the new biofuel combustion demand values based on the condition values. A total of 64 days of output data were generated, with each day's data containing 96 data points and 50 sets of scenario data generated each day. Three of the days of scenario data are selected to compare with the real data curve, which are the 28th, 30th, and 60th days of data, and the results of data generation are shown in Fig. 4, with Figs. (a), (b), and (c) representing the 28th, 30th, and 60th days, respectively.

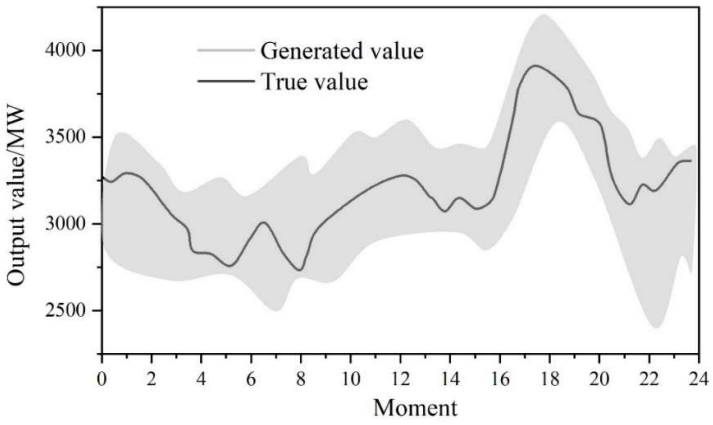
As can be seen from the data images in Fig. 4, the trend of the generated data set is consistent with the trend of the actual output value of the power plant, and the interval range formed by the generated 50 sets of data can contain the real data in a more complete way, with a high data coverage, and the error between the generated value and the real value is in the range of ±250-±300. From the above result process, it can be seen that the Wasserstein WGAN model proposed in this paper can accurately learn the relationship between the conditional value and the target value, and then generate the determined scene data based on the conditional value.



(a)The scene is generated on day 28



(b)The scene is generated on day 30

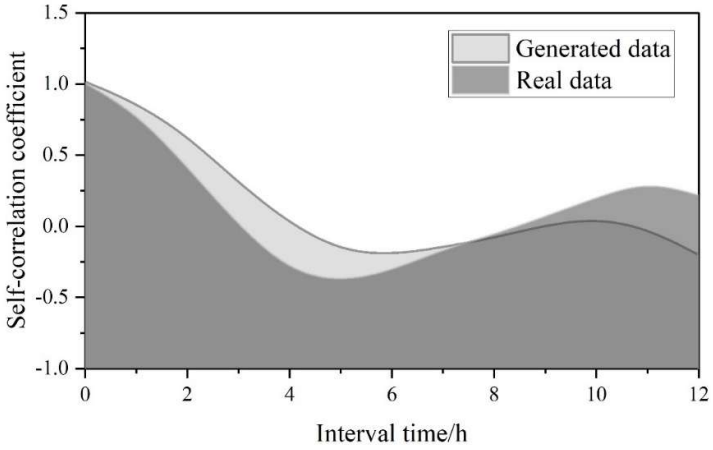


(c)The scene is generated on day 60

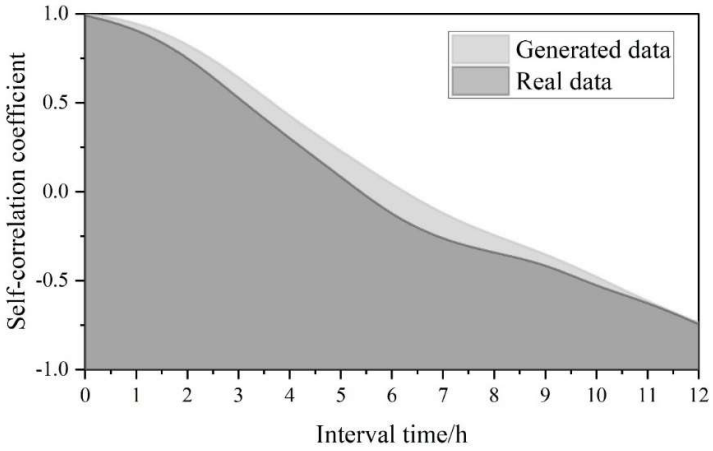
Fig. 4. The power generation results

2.3.3. Autocorrelation coefficient. In addition, in order to verify that the generated data have temporal correlation, ACF is used for verification. The average values of 50 groups of generated data and real data on the 28th, 30th, and 60th days in the test set are selected as samples, and the ACF of the generated scenes is calculated with the time interval unit of 15 min. The autocorrelation coefficients are schematically shown in Fig. 5, and Figs. (a), (b), and (c) are for the 28th, 30th, and 60th days. From the dark gray area in Fig. 5, it can be seen that the real scene sequence has autocorrelation

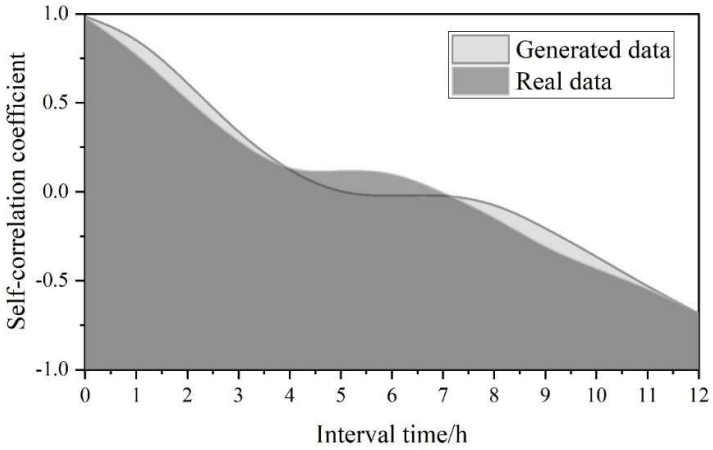
in the first 3~7h of the interval, and then the autocorrelation decreases with the increase of the time interval, which indicates that the output data have strong longitudinal time correlation, and confirms the correctness of the longitudinal splicing of the noise and the conditions in the process of data input. The autocorrelation coefficients of the generated data are shown in the light gray area in the figure, which shows that the autocorrelation coefficients of the generated data fit the autocorrelation coefficient trend of the real data and have high fitting accuracy. This indicates that the data generated by the method in this paper can reflect the time series correlation of the real data.



(a)Day 28 autocorrelation coefficient



(b)Day 30 autocorrelation coefficient



(c)Day 60 autocorrelation coefficient

Fig. 5. Auto correlation coefficient

2.3.4. Comparison of results of different methods. In order to verify the effectiveness of the proposed output prediction model, the model in this paper is compared and verified with the CGAN model to analyze the advantages and disadvantages between the 2 models. The data of 320 days, 480 days and 640 days are selected to train the CGAN model, but no condition value is added to the input data, and the same loss function training method as Wasserstein WGAN is used to obtain the mapping relationship between the data. After completing the data generation the results are compared using statistical learning metrics for all 64 days of test data and the results obtained are shown in Table 1.

From the results in Table 1, it can be seen that:

1) with the increase in the number of training days, the MAE and RMSE indicators of the force data generated by the traditional method CGAN model are larger and show fluctuations, which indicates that the data generation accuracy of the traditional method is not high and the generation effect is not stable. On the other hand, the WGAN model based on Wasserstein distance in this paper, the MAE and RMSE indexes of the generated data are smaller, which indicates that the data generated by the method in this paper are closer to the real data, and the accuracy of scenario generation is higher. Taking 640 sample days as an example, the MAE and RMSE values of this paper's model are 315.595 and 397.485, respectively, while the MAE and RMSE values of the CGAN model are 582.169 and 664.486.

2) Taking 640 days as an example, in terms of mean and standard deviation, the difference between the mean and standard deviation of this paper's model and the real value are 188.97 and 116.64, respectively, while the difference between the mean and standard deviation of the CGAN model and the real value are 286.859 and 542.627. Compared with CGAN, the Wasserstein WGAN model proposed in this paper generates data of the proposed Wasserstein WGAN model are closer to the relevant indicator values of the real data. And from the standard deviation of the data generated by the two types of models, it can be seen that the standard deviation of the data generated by the traditional method CGAN and the standard deviation of the real output data have a large difference, which does not reflect the volatility of the real data itself. In contrast, the standard deviation of the data generated by the Wasserstein WGAN proposed in this paper is closer to the standard deviation of the real output data, which indicates that the data generated by this paper's method can better express the volatility of the real new biofuel combustion data, and can more effectively characterize the uncertainty of the real new biofuel combustion output data.

Table 1. Evaluation indicators for the different models

Sample days	Model	Evaluation index				
		MAE	RMSE	R^2	Average value /M /MW	Standard deviation /MW
Real data e		-	-	-	3318.455	619.485
640 days	Textual model	315.595	397.485	0.9152	3129.485	502.845
	CGAN	582.169	664.486	0.7512	3031.596	76.858
480 days	Textual model	334.453	416.985	0.8154	3201.445	498.485
	CGAN	541.569	624.895	0.7054	3186.641	94.552
320 days	Textual model	379.699	470.954	0.9569	3108.945	585.964

	CGAN	576.894	654.974	0.8054	3054.545	108.455

3. Scenario design of new biofuel combustion using WGAN-GP modeling

3.1. Biofuels and combustion characteristics

Biofuel is a renewable energy source, which is mainly obtained from the conversion of renewable biomass materials. The biomass materials mainly include starch, waste straw, algae, oil and grease. The combustion characteristics of biofuels include combustion efficiency, oil-poor flameout and ignition performance, etc. There have been a large number of studies on the combustion characteristics of biofuels, but most of these studies have focused on the types of biofuels, purification, etc., and seldom investigated the behavior of biofuels. However, before the study of combustion behavior of new biofuels, baking pretreatment of the fuel is needed. Roasting pretreatment can improve the physicochemical properties of solid biofuels, and in order to clarify the effect of the roasting process on the utilization of biofuels in thermal conversion, it is necessary to study the combustion behavior of roasted biochar.

3.2. Catalytic synthesis of EMF technology

It is well known that acidic catalysts are the key to the synthesis of EMF from converted biomass sugar, so the search for efficient and practical catalytic systems has become the primary task and research focus of this reaction technology, which is one of the core technologies for successful industrialization. The currently developed catalysts for the synthesis of EMF mainly include inorganic liquid acids, metal salts, molecular sieves, sulfonic acid-based solid acids, functionalized ionic liquids, and heteropolyacids. In the following, the catalytic systems are divided into homogeneous catalytic system and multiphase catalytic system according to the characteristics of the reaction process of EMF synthesis to summarize separately.

3.3. 5-Ethoxymethylfurfural (EMF)

3.3.1. Properties and applications of EMFs. EMF is produced by the conversion reaction of carbohydrates or lignocellulose and can be used as a petroleum substitute. In engine tests, blending 17% EMF into commercial diesel significantly reduced carbon smoke emissions by 16%. EMF has a high energy density (30.3 MJ/L), comparable to that of diesel fuel (33.6 MJ/L), a high boiling point (508 K), a low flash point (~110°C), high oxidative stability, and low toxicity. Due to its low turbidity point, EMF can also greatly improve the cold flow characteristics of biodiesel. In addition to this, EMF is widely used as a plasticizer in the manufacture of fragrances, additives, fine chemicals and pharmaceuticals.

3.3.2. Development trend of EMF preparation. EMF can be synthesized from a variety of carbohydrates (e.g. fructose, glucose, cellulose, inulin and straw) in acidic ethanol solutions. It has been shown that high EMF yields (more than 90%) can be easily obtained using fructose or HMF as feedstock, however, EMF yields obtained by direct conversion of lignocellulose are less than 20%. In the alcoholic conversion of glucose, solvent and catalyst are the two basic factors affecting product selectivity and EMF yield.

1) Solvent system. Solvent effect is one of the important factors affecting the catalytic process of biomass. Solvent effect is a common phenomenon in homogeneous and multiphase catalysis, which may affect the solubility, stability and reaction rate of the reaction system. Since various properties of solvents affect catalytic reactions in many different ways, solvent effects are a complex problem involving solubility, mass transfer, interaction with initial substances or products, transition state

stabilization, and interaction with catalysts. Among these aspects, the interaction between solvent and catalyst is the most important factor.

Addition of organic solvents to the reaction medium improves catalytic activity and selectivity, and organic solvents can modulate the decomposition reactions of sugars to increase the yield of value-added chemicals. It has been found that a 27% yield of EMF from glucose can be obtained in a solvent system of ethanol and dimethylsulfoxide (DMSO) using a sulfonic acid-based solid acid as a catalyst. It has been shown that in the reaction system with solvent, the EMF yield from fructose was 8.13% higher than that without solvent. Scholars speculate that the organic solvent may provide a hydrophobic environment that contributes to the EMF yield.

2) Catalytic system. Homogeneous acid catalysts have attracted the interest of many researchers due to their effective catalysis and low cost. Some scholars catalyzed glucose with a very low concentration of sulfuric acid in methanol solution and finally obtained 50% yield of methyl acetyl propionate. Some scholars also catalyzed wheat straw with sulfuric acid and obtained 17.91% EL yield. Aromatic hydrocarbons of 20% were obtained. However, reuse of homogeneous inorganic acid catalysts leads to equipment corrosion and recycling difficulties.

3.4. Preparation of EMF in a novel biofuel generation scenario

3.4.1. Glucosolysis experiments. In the reaction to test the ultra-low acid-catalyzed glucose for the production of EMF reaction intermediates, a certain mass of glucose was weighed out using an electronic balance, and 80 mL of anhydrous ethanol and 0.1 wt% of sulfuric acid were measured and placed into a 150 mL autoclave. In the USY-catalyzed glucose alcoholysis reaction, a certain amount of glucose and USY were added to the reactor and then, 60 mL of solvent was added while stirring magnetically. The stirring speed during heating was 250 rpm. Start timing when the autoclave is heated to a predetermined temperature and adjust the rotation speed to 500 rpm. At the end of the reaction, the autoclave was cooled to room temperature in a cold water bath. The reaction solution was then filtered and diluted to a certain power for the next experiment. Each set of experiments was repeated three times and the average value was taken. The yields used in the experiments were all molar yields.

3.4.2. Analytical methods. EMF and HMF were detected by high performance liquid chromatography (HPLC) equipped with a UV detector (280 nm) and a ZORBAX Eclipse XDB-C18 column (4.6×250 mm, 5 μm). The column temperature was maintained at 30 ℃, and the mobile phase was acetonitrile/0.1% aqueous acetic acid (30/70 v/v) at a flow rate of 0.6 mL min-1. The detection wavelength was 280 nm, and the injection volume was 5 μL. The EL was separated and detected using gas chromatography (GC) equipped with a HP-5MS capillary column and a FID detector. The starting temperature was 60 ℃, and was increased to 280 ℃ at a 10 ℃ min-1 rate to 280 °C, and then held at 280 °C for 2 min. During this process, the flow rate of the carrier gas (N_2) was kept at $1.0mL\,\min^{-1}$, and the temperatures of the injector and detector were 240 °C and 250 °C, respectively. In addition, the distribution of the liquid products was confirmed by gas chromatography-mass spectrometry (GC/MS). An HP-5MS capillary column (30 m × 0.32 mm × 0.50 um) was used. High purity helium was used as the carrier gas at a flow rate of $1.6mL\,\min^{-1}$. The temperature program started at 40 °C and was heated to 220 °C at a rate of 10 °C per minute and held for 8 min. The mass spectrometry detector was operated in electron ionization mode (70 eV, 50-400 m/z) at 220 °C.

3.5. Catalyst Reuse and Characterization

3.5.1. USY Reuse Experiments. Reuse experiments of USY were carried out at a catalyst loading of 1.8 wt% and a THF volume fraction of 30%. Fresh USY was recorded as R0. After each reaction, the

catalyst was filtered from the reaction mixture and dried at 105 °C for 24 h. The recovered catalyst was then calcined in a muffle furnace at 500 °C for 4 h and reused in the next cycle.

3.5.2. Recovery of USY characterization methods. Scanning electron microscope (SEM) images of USY were obtained on a Zeiss Sigma 500. The working distance was 10.5 mm, the operating power was 5.00 KW, and the EHT (ultrahigh pressure) was 3.00 kv.The adsorption-desorption data of N2 were recorded on a Micromeritics APSP 2460 device after the samples were degassed for 5 h at 150 °C.The specific surface area of these samples was calculated using a BET model.The acidic sites of USY were desorbed by temperature-programmed ammonia (NH3-TPD) was evaluated on a TP-5080-B equipped with a TCD detector. FT-IR spectra of adsorbed pyridine were utilized to identify the type of acid site of USY on a Bruker Tensor 27. Pyridine vapor was then adsorbed in the sample for 30 min at room temperature and degassed at 200°C. The IR spectra were then recorded.

3.6. Results and Discussion

3.6.1. Structural characterization of catalysts. The synthesis experiments of EMF were carried out under the new biofuel outgrowth scenario generated by using the WGAN-GP model, and its chemical structure was characterized by Fourier infrared spectroscopy analysis of the material as shown in Fig. 6, and the chemical structure of the material was characterized by Fourier infrared spectroscopy analysis of the material as shown in Fig. 6.The PTFE@HCMSs denotes the hydrothermal-treated samples, and the PTFE@ACMS is the annealed samples, and the final sulfonated obtained sample is PTFE@ACMS-SO_3H. The broad peaks in the range of 3000-3329 cm^{-1} originate from the stretching vibration of the conjugated O-H, the peak at 2936 cm^{-1} is caused by the stretching vibration of C-H, and the absorption peaks at 1324 cm^{-1}, 1159 cm^{-1} and 859 cm^{-1} are the stretching vibration peaks of C=O, C=C and C-O, respectively. By comparing the curves of the samples treated under different conditions, it was found that the strong absorption peaks at 759 cm^{-1} and 54 cm^{-1} for the PTFE@ACMS-SO_3H samples were due to the asymmetric telescopic vibration of O=S=O in -SO_3H, which indicated that the acidic group -SO_3H was successfully loaded on the surface of the material. The weakening of the C-H stretching vibration absorption peaks after sulfonation of the material is also attributed to the formation of O=S=O.

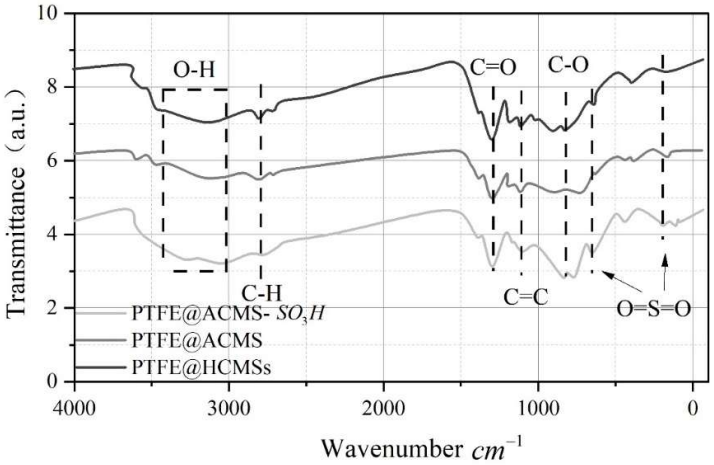
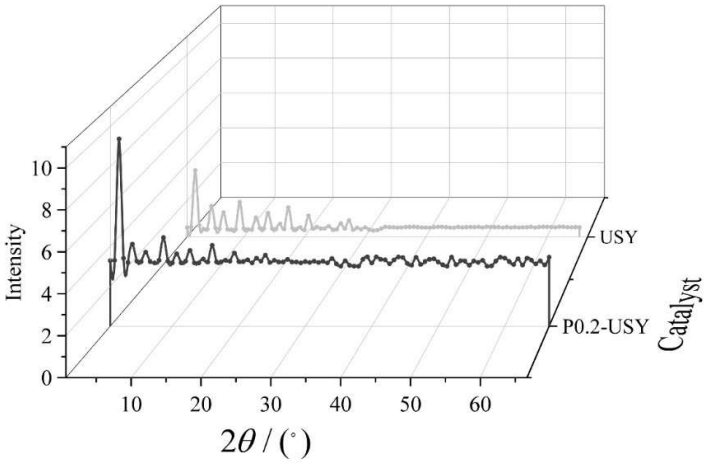


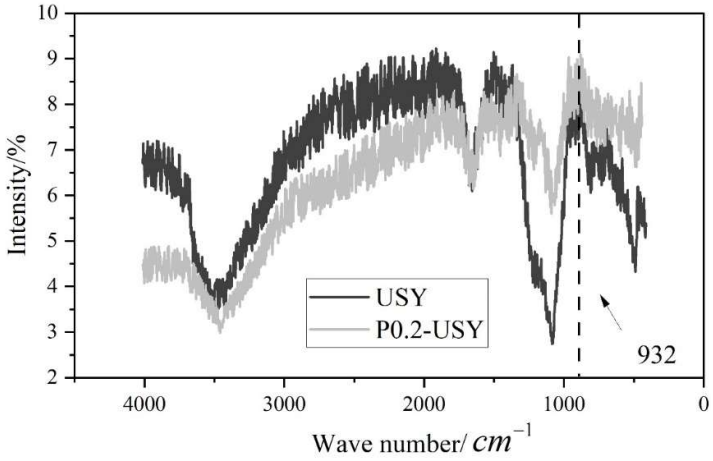
Fig. 6. FTIR diagram of the materials

3.6.2. Catalyst Modification and Reuse Characterisation. USY and P0.2-USY were characterised to analyse the effects of modification on the structure, functional groups, specific surface area and acid amount of USY. Figure 7 shows the structural and functional group changes of the USY catalysts before and after modification, with X-ray diffraction in Fig. (a) and Fourier infrared spectral transformation

in Fig. (b). The characteristic diffraction peaks of the catalysts before and after modification appeared at 2θ of 10.26°, 11.86°, 15.53°, 18.86°, 20.36°, 23.56° and 27.34°. The peak heights of the small-angle peaks changed after modification, and the peak at 2θ of 10.26° showed a large change in peak height. There is a characteristic peak of 1 surface hydroxyl group between 932 cm^{-1} for USY before and after modification.



(a)X-ray diffraction



(b)Fourier transform infrared reflection

Fig. 7. The modification of usy catalyst and the change of the functional group

4. Conclusion

In this paper, based on generative adversarial networks, Wasserstein distance is introduced and weight trimming is replaced by gradient penalty, and the structure of WGAN-GP discriminator is improved by a self-attentive model to generate high-quality data samples more efficiently and stably. Using the model to generate new biofuel combustion scenario data. The model is trained and its performance is analysed by simulation experiments to simulate a novel biofuel combustion scenario.

1) The model designed in this paper in the process of continuous iteration generator generated data in the process of continuous approximation of the real data, in the number of iterations for the beginning of 100 times, the loss function of the generator and the discriminator gradually converge to stability, and in the axis of the 0 value of the vicinity of the Nash equilibrium is reached. Therefore, the Wasserstein WGAN-GP loss function model with penalty term proposed in this paper has better stability as well as convergence than the traditional loss function model without penalty.

2) Comparing the generated values and the true values of this paper's model, the error between the generated values and the true values is between ±250 and ±300, and the Wasserstein WGAN model proposed in this paper can accurately learn the relationship between the conditional values and the target values, and then generate the determined scene data based on the conditional values.
3) The synthesis experiments of EMF were performed under the combustion scenario of the new biofuel generated by the WGAN-GP model, and the characterisation of the new biofuel and its chemical structure were analysed using Fourier IR spectroscopy. The peak heights at small angles appeared to change after modification, and the peak at 10.26° for 2θ showed a large change in peak height. There is a characteristic peak of 1 surface hydroxyl group between 932 cm^{-1} for USY before and after modification.

References

[1] Ambaye, T. G., Vaccari, M., Bonilla-Petriciolet, A., Prasad, S., van Hullebusch, E. D., & Rtimi, S. (2021). Emerging technologies for biofuel production: A critical review on recent progress, challenges and perspectives. Journal of Environmental Management, 290, 112627.

[2] Kour, D., Rana, K. L., Yadav, N., Yadav, A. N., Rastegari, A. A., Singh, C., ... & Saxena, A. K. (2019). Technologies for biofuel production: current development, challenges, and future prospects. Prospects of renewable bioprocessing in future energy systems, 1-50.

[3] Binod, P., Gnansounou, E., Sindhu, R., & Pandey, A. (2019). Enzymes for second generation biofuels: recent developments and future perspectives. Bioresource Technology Reports, 5, 317-325.

[4] Bhatia, S. K., Kim, S. H., Yoon, J. J., & Yang, Y. H. (2017). Current status and strategies for second generation biofuel production using microbial systems. Energy Conversion and Management, 148, 1142-1156.

[5] Liu, Y., Cruz-Morales, P., Zargar, A., Belcher, M. S., Pang, B., Englund, E., ... & Keasling, J. D. (2021). Biofuels for a sustainable future. Cell, 184(6), 1636-1647.

[6] Adeniyi, O. M., Azimov, U., & Burluka, A. (2018). Algae biofuel: current status and future applications. Renewable and sustainable energy reviews, 90, 316-335.

[7] Robertson, G. P., Hamilton, S. K., Barham, B. L., Dale, B. E., Izaurralde, R. C., Jackson, R. D., ... & Tiedje, J. M. (2017). Cellulosic biofuel contributions to a sustainable energy future: Choices and outcomes. Science, 356(6345), eaal2324.

[8] Lin, C. Y., & Lu, C. (2021). Development perspectives of promising lignocellulose feedstocks for production of advanced generation biofuels: A review. Renewable and Sustainable Energy Reviews, 136, 110445.

[9] Oh, Y. K., Hwang, K. R., Kim, C., Kim, J. R., & Lee, J. S. (2018). Recent developments and key barriers to advanced biofuels: A short review. Bioresource Technology, 257, 320-333.

[10] Prussi, M., O'connell, A., & Lonza, L. (2019). Analysis of current aviation biofuel technical production potential in EU28. Biomass and Bioenergy, 130, 105371.

[11] ElFar, O. A., Chang, C. K., Leong, H. Y., Peter, A. P., Chew, K. W., & Show, P. L. (2021). Prospects of Industry 5.0 in algae: Customization of production and new advance technology for clean bioenergy generation. Energy Conversion and Management: X, 10, 100048.

[12] Peskett, L., Slater, R., Stevens, C., & Dufey, A. (2022). Biofuels, agriculture and poverty reduction. ODI.

[13] Mahlia, T. M. I., Syazmi, Z. A. H. S., Mofijur, M., Abas, A. P., Bilad, M. R., Ong, H. C., & Silitonga, A. S. (2020). Patent landscape review on biodiesel production: Technology updates. Renewable and Sustainable Energy Reviews, 118, 109526.

[14] Agarwal, A. K., Gupta, J. G., & Dhar, A. (2017). Potential and challenges for large-scale application of

biodiesel in automotive sector. Progress in Energy and Combustion Science, 61, 113-149.

[15] Correa, D. F., Beyer, H. L., Fargione, J. E., Hill, J. D., Possingham, H. P., Thomas-Hall, S. R., & Schenk, P. M. (2019). Towards the implementation of sustainable biofuel production systems. Renewable and Sustainable Energy Reviews, 107, 250-263.

[16] Bashir, M. A., Wu, S., Zhu, J., Krosuri, A., Khan, M. U., & Aka, R. J. N. (2022). Recent development of advanced processing technologies for biodiesel production: A critical review. Fuel Processing Technology, 227, 107120.

[17] Yaegashi, J., Kirby, J., Ito, M., Sun, J., Dutta, T., Mirsiaghi, M., ... & Gladden, J. M. (2017). Rhodosporidium toruloides: a new platform organism for conversion of lignocellulose into terpene biofuels and bioproducts. Biotechnology for biofuels, 10, 1-13.

[18] Araújo, K., Mahajan, D., Kerr, R., & da Silva, M. (2017). Global biofuels at the crossroads: an overview of technical, policy, and investment complexities in the sustainability of biofuel development. Agriculture, 7(4), 32.

[19] Mendonça, I. M., Paes, O. A., Maia, P. J., Souza, M. P., Almeida, R. A., Silva, C. C., ... & de Freitas, F. A. (2019). New heterogeneous catalyst for biodiesel production from waste tucumã peels (Astrocaryum aculeatum Meyer): Parameters optimization study. Renewable Energy, 130, 103-110.

[20] Neuling, U., & Kaltschmitt, M. (2017). Review of biofuel production-feedstock, processes and markets. J Oil Palm Res, 29(2), 137-167.

[21] Zabaniotou, A. (2018). Redesigning a bioenergy sector in EU in the transition to circular waste-based Bioeconomy-A multidisciplinary review. Journal of Cleaner Production, 177, 197-206.

[22] Saravanan, A. P., Pugazhendhi, A., & Mathimani, T. (2020). A comprehensive assessment of biofuel policies in the BRICS nations: Implementation, blending target and gaps. Fuel, 272, 117635.

[23] Alalwan, H. A., Alminshid, A. H., & Aljaafari, H. A. (2019). Promising evolution of biofuel generations. Subject review. Renewable Energy Focus, 28, 127-139.

[24] Li, P., Sakuragi, K., & Makino, H. (2019). Extraction techniques in sustainable biofuel production: A concise review. Fuel Processing Technology, 193, 295-303.

[25] Saravanan, A., Murugan, M., Reddy, M. S., & Parida, S. (2020). Performance and emission characteristics of variable compression ratio CI engine fueled with dual biodiesel blends of Rapeseed and Mahua. Fuel, 263, 116751.

[26] Ganesan, R., Manigandan, S., Samuel, M. S., Shanmuganathan, R., Brindhadevi, K., Chi, N. T. L., ... & Pugazhendhi, A. (2020). A review on prospective production of biofuel from microalgae. Biotechnology Reports, 27, e00509.

[27] Souza, G. M., Ballester, M. V. R., de Brito Cruz, C. H., Chum, H., Dale, B., Dale, V. H., ... & Van der Wielen, L. (2017). The role of bioenergy in a climate-changing world. Environmental development, 23, 57-64.

[28] Rodionova, M. V., Poudyal, R. S., Tiwari, I., Voloshin, R. A., Zharmukhamedov, S. K., Nam, H. G., ... & Allakhverdiev, S. I. (2017). Biofuel production: challenges and opportunities. International Journal of Hydrogen Energy, 42(12), 8450-8461.

[29] Joshi, G., Pandey, J. K., Rana, S., & Rawat, D. S. (2017). Challenges and opportunities for the application of biofuel. Renewable and Sustainable Energy Reviews, 79, 850-866.

[30] Hossain, N., Mahlia, T. M. I., & Saidur, R. (2019). Latest development in microalgae-biofuel production with nano-additives. Biotechnology for biofuels, 12, 1-16.

[31] Raud, M., Kikas, T., Sippula, O., & Shurpali, N. J. (2019). Potentials and challenges in lignocellulosic biofuel

production technology. Renewable and Sustainable Energy Reviews, 111, 44-56.

[32] Rathour, R. K., Ahuja, V., Bhatia, R. K., & Bhatt, A. K. (2018). Biobutanol: New era of biofuels. International Journal of Energy Research, 42(15), 4532-4545.

[33] Klein, B. C., Chagas, M. F., Watanabe, M. D. B., Bonomi, A., & Maciel Filho, R. (2019). Low carbon biofuels and the New Brazilian National Biofuel Policy (RenovaBio): a case study for sugarcane mills and integrated sugarcane-microalgae biorefineries. Renewable and Sustainable Energy Reviews, 115, 109365.

[34] Grassi, M. C. B., & Pereira, G. A. G. (2019). Energy-cane and RenovaBio: Brazilian vectors to boost the development of Biofuels. Industrial crops and products, 129, 201-205.

[35] Jafarihaghighi, F., Ardjmand, M., Salar Hassani, M., Mirzajanzadeh, M., & Bahrami, H. (2020). Effect of fatty acid profiles and molecular structures of nine new source of biodiesel on combustion and emission. ACS omega, 5(26), 16053-16063.

[36] Shuyan Hao,Ting Xia,Ruizhi Zhang & Meng Guo. (2024). Clustering Cu-S based compounds using periodic table representation and compositional Wasserstein distance. Scientific Reports(1),31602-31602.

[37] Qiao Ning & Zedong Qi. (2025). WGAN-GPGlu: A semi-supervised model based on double generator-Wasserstein GAN with gradient penalty algorithm for glutarylation site identification. Computers in Biology and Medicine109328-109328.

[38] Junjie Liang,Renjie Liang & Dongxia Wang. (2025). A novel lightweight model for tea disease classification based on feature reuse and channel focus attention mechanism. Engineering Science and Technology, an International Journal101940-101940.

[39] Yingjie Qiao & Lizhi Xing. (2024). Deep Learning Based Oracle Style Migration and Generation Techniques. IAENG International Journal of Computer Science(11),

[40] Artzai Picon,Daniel Mugica,Itziar Eguskiza,Arantza Bereciartua Perez,Javier Romero,Carlos Javier Jimenez... & Ramon Navarra Mestre. (2025). Taxonomic hierarchical loss function for enhanced crop and weed phenotyping in multi-task semantic segmentation. Smart Agricultural Technology100761-100761.