

Article

Extreme Wind Power Output Scenario Generation Method Guided and Constrained by Statistical Features

Dan Li ^{1,2,*}, Xiangyang Liang ^{1,2,*}, Minghan Qu ^{1,3}, Yawen Zhen ^{1,3}, Zhaoxi Lin ^{1,3} and Bin Yao ^{1,3}

¹ College of Electrical and New Energy, China Three Gorges University, Yichang 443002, China; lucy2140@163.com (D.L.); minghan0921@163.com (M.Q.); zzino01@163.com (Y.Z.); linzhaoxi14@outlook.com (Z.L.); 18968172977@163.com (B.Y.)

² Hubei Provincial Key Laboratory for Operation and Control of Cascaded Hydropower Station, China Three Gorges University, Yichang 443002, China

³ Hubei Provincial Collaborative Innovation Center for New Energy Microgrid, China Three Gorges University, Yichang 443002, China

* Correspondence: 19086826587@163.com

Abstract

The increasing penetration of renewable energy and the frequent occurrence of extreme weather events have significantly heightened the uncertainty in power system operations. Simultaneously, the scarcity of renewable energy output samples under extreme meteorological conditions constrains the accurate assessment of extreme risks in system planning and dispatch. To bridge this gap, this work aims to propose a method for generating extreme wind power output scenarios that possess both diversity and statistical accuracy under limited sample conditions. To address this, this paper proposes a method for generating scenarios of extreme wind power output guided and constrained by statistical features. First, multidimensional statistical features are extracted from historical wind power output scenarios and combined, and a quantile threshold method is applied to screen out extreme wind power output scenarios. Subsequently, based on differentiated application requirements of the power system, extreme scenarios undergo preliminary classification followed by category-specific clustering analysis, achieving refined classification of the scenario set. Building on this, an improved generative adversarial network model is constructed, and the Wasserstein distance and gradient penalty mechanism are introduced to enhance training stability. Additionally, a statistical feature self-attention mechanism and feature loss function are designed to effectively constrain the consistency between generated scenarios and real scenarios in key statistical features. Results demonstrate that the proposed method can generate a set of extreme wind power output scenarios with both diversity and statistical accuracy under limited sample conditions, providing effective data support for system safety operation and risk prevention and control.

Keywords: scenario generation; wind power output; statistical features; feature constraint; generative adversarial network

Academic Editors: Galih Bangga,
Martin Otto Laver Hansen and
Tapas Mallick

Received: 3 January 2026

Revised: 29 January 2026

Accepted: 7 February 2026

Published: 14 February 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

To effectively address the severe challenges posed by the energy crisis and environmental pollution, the nation has explicitly set forth the strategic goals of “carbon peak and carbon neutrality,” focusing on promoting the large-scale replacement of traditional fossil fuels with renewable energy sources such as wind power and photovoltaics [1,2]. In this

context, the grid integration ratio of renewable energy sources like wind power within the new power system is rapidly increasing. The inherent variability and intermittency of renewable energy sources like wind power, compounded by the frequent occurrence of extreme weather events in recent years, have significantly increased the uncertainty of their output; this poses severe challenges to the safe and stable operation of power systems and their economically efficient dispatch [3]. Power system planning and operations heavily rely on data-driven methods (such as probabilistic forecasting and stochastic optimization) to address these uncertainties, and the effectiveness of these methods largely depends on the quality of historical data [4]. However, historical output samples under extreme weather conditions are often scarce, and this scarcity makes it difficult for the system to accurately assess the tail risks triggered by extreme events during power forecasting, medium and long-term planning, and real-time dispatch decisions, thereby severely limiting the reliability and robustness of operational strategies [5]. Therefore, to address the challenge of scarce historical samples under extreme weather conditions, there is an urgent need to develop a scenario generation method that can integrate the characteristics of extreme wind power output, thereby providing effective data support for the system to cope with extreme events.

Currently, scenario generation methods are primarily categorized into two types: statistical probability modeling methods and deep learning methods based on big data [6]. Probability modeling methods fit historical data distributions (such as normal or Weibull distributions) or employ kernel density estimation and copula functions to construct joint distributions, which are then used to generate random scenarios through Monte Carlo simulation or Latin hypercube sampling [7–9]. However, this method relies on the assumptions of sufficient data and a stationary distribution. Extreme wind power output events exhibit sparse samples and dynamic instability, making it challenging for traditional methods to accurately model their probability distributions. Therefore, there is an urgent need to explore deep generative methods capable of directly learning complex distribution characteristics from limited extreme samples.

With the development of artificial intelligence technology, generative models represented by generative adversarial networks (GANs) [10–12] and variational auto-encoders (VAEs) [13] have been widely applied in renewable energy output scenario generation due to their exceptional capabilities in feature extraction and complex distribution fitting. Existing research has effectively enhanced training stability and generative controllability under conventional data by introducing mechanisms such as conditional generation [10,11], Wasserstein distance and gradient penalties [12,13]. However, improved training stability has not resolved the core challenge of generating accurate results in extreme scenarios. This stems primarily from network architectures that lack targeted design for the statistical features of extreme scenarios, resulting in insufficient capture of critical extreme features and ultimately limiting the accuracy of generated outcomes.

Despite these limitations, some scholars have begun exploring methods for generating extreme scenarios for renewable energy, achieving preliminary progress. Reference [14] defines different extreme metrics based on wind power features and employs conditional generative adversarial networks to achieve controllable generation of extreme scenarios. Reference [15] constructs an annual time series model capable of integrating multiple short-timescale extreme meteorological events, generating continuous meteorological scenarios and mapping them to renewable energy output. Reference [16] considers the influence of midterm weather processes, employing a conditional diffusion model for the controllable generation of extreme wind power scenarios. These studies provide valuable insights for generating extreme scenarios for renewable energy. However, most existing methods are developed under relatively ideal conditions with comprehensive data, and their scenario definitions often rely on detailed meteorological data. In practical

applications, severe historical data shortages or difficulties in obtaining high-precision meteorological data are common challenges, limiting the applicability of existing approaches.

Moreover, as extreme weather events become more frequent, their impact on power systems grows increasingly multifaceted, while different application scenarios prioritize distinct aspects of extreme output characteristics. For instance, peak shaving capability assessments prioritize sustained low output scenarios [17,18], ramping risk warnings require capturing short-term high ramping rates and fluctuation features, while reserve capacity allocation focuses on medium and long-term high ramping rates and peak–valley difference scenarios [19,20]. Existing research predominantly focuses on single features, lacking a scenario generation framework capable of systematically distinguishing and generating sets tailored to differentiated needs such as peak shaving capability assessment, ramping risk warning, and reserve capacity allocation. Therefore, it is necessary to refine the classification of extreme scenario sets based on the practical characteristics of power system applications, generating highly targeted extreme scenario collections to support the practical demands of power system operation and planning.

Based on the above research, this paper proposes a framework for generating extreme scenarios using historical wind power output features, addressing the constraint of limited extreme wind power output data. Key advantages and innovations include:

- (1) Utilizing historical wind power output features, including daily average, peak–valley difference, maximum ramp rate, and fluctuation frequency. The framework employs a quantile threshold method to screen extreme wind power output scenarios. This achieves scenario identification independent of meteorological data while efficiently leveraging scarce extreme samples.
- (2) By integrating differentiated application requirements of power systems, extreme scenarios undergo preliminary classification followed by refined categorization within each class. This generates customized scenario sets for diverse application requirements.
- (3) Extreme scenarios are generated using traditional generative adversarial networks, incorporating Wasserstein distance constraints and gradient penalty mechanisms to enhance training stability. Statistical feature self-attention guidance and statistical feature loss function constraints are employed to improve consistency between generated and real scenarios in key statistical features.

2. Model Structure and Framework

2.1. Analysis of Wind Power Output Features Under Extreme Weather Conditions

Extreme meteorological conditions affecting the safe and stable operation of power grids encompass not only rare, high-intensity events defined in meteorology (such as torrential downpours, super typhoons, and extreme cold waves), but also weather phenomena capable of inducing sharp fluctuations in renewable energy output. Under such conditions, the output from renewable sources like wind and solar power may exhibit complex variations characterized by “sudden surges like waves, abrupt drops like cliffs, and fluctuations like tides,” posing severe challenges to the stable operation of power systems [21,22]. To address this challenge, it is necessary to conduct an analysis of wind power output features under extreme weather conditions.

Meteorological factors affecting wind power output are diverse, primarily including wind speed, wind direction, air temperature, precipitation, and ice accumulation. Among these, wind speed exerts the most significant influence, directly determining the cut-in, rated and cut-out operating states of the wind turbine. Meanwhile, meteorological conditions such as air temperature and ice accumulation can substantially alter the

aerodynamic characteristics and mechanical performance of wind turbines, thereby constraining their actual power generation capacity. Representative weather conditions that typically limit wind power output are shown in Figure 1, including high winds, low temperatures, and snowfall.

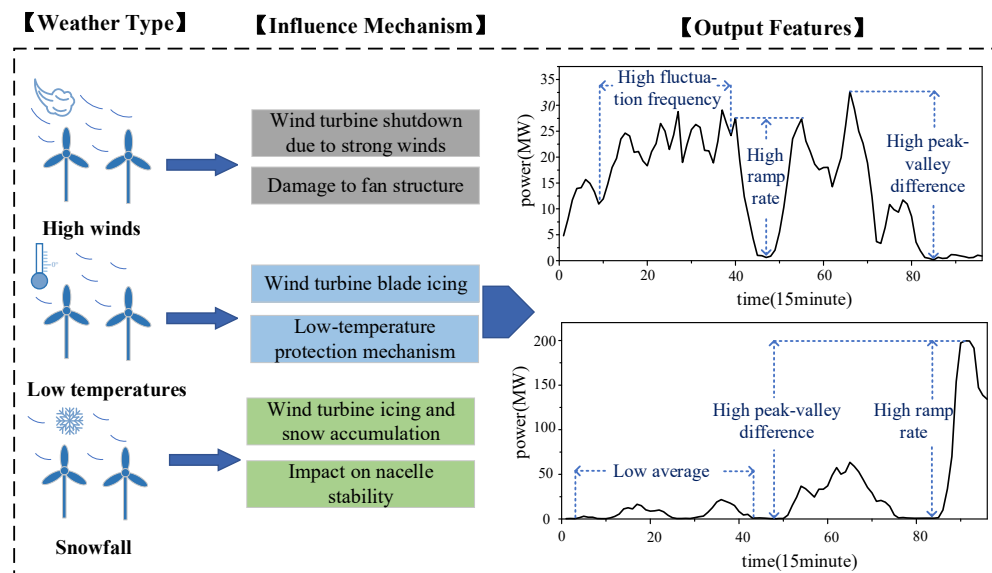


Figure 1. Analysis of typical extreme weather events.

The aforementioned extreme weather conditions, represented by high winds, low temperatures, and snowfall, though differing in meteorological causes, all induce complex features in wind power output, such as high volatility and steep ramping. To quantitatively analyze these features, this paper utilizes historical wind power output data and combines statistical and time series analysis theories to select the following four key indicators for characterization. These indicators are then used for subsequent extreme scenario screening:

- (1) Daily average

$$P_{avg} = \frac{1}{N} \sum_{t=1}^N P_t \quad (1)$$

Characterizes the level of daily power generation capacity. Here, P_t represents the power value at the t th sampling point, and N denotes the total number of daily samples.

- (2) Peak–valley difference

$$\Delta P = P_{max} - P_{min} \quad (2)$$

Reflects the fluctuation range of daily output. P_{max} and P_{min} represent the daily peak and trough power values, respectively.

- (3) Maximum ramp rate

$$R_{ramp} = \max \left(\left| \frac{P_{t2} - P_{t1}}{\Delta T} \right| \right) \quad (3)$$

Describes the rate of change in power [23]. Takes the maximum absolute value of the power change over the ΔT sampling interval.

- (4) Fluctuation frequency

$$N_{fluc} = \sum_{t=2}^{N-1} I[\text{sgn}(P_t - P_{t-1}) \neq \text{sgn}(P_{t+1} - P_t)] \quad (4)$$

Used to quantify the number of fluctuations in the wind power output curve during the observation period. Here, $I(\cdot)$ is an indicator function (assigning 1 when the condition holds, 0 otherwise), and $\text{sgn}(\cdot)$ is a sign function that extracts the direction of output changes.

To construct a dataset of extreme wind power output scenarios, this study employs a quantile threshold method for scenario selection based on the aforementioned four key statistical indicators. This approach sets multi-level quantile thresholds for each indicator and combines these thresholds to identify extreme wind power output scenarios from historical data that simultaneously exhibit multiple extreme features.

2.2. Application Analysis of Extreme Wind Power Output Scenarios in Power Systems

Based on the set of extreme wind power output scenarios, to meet the differentiated requirements of various power system application scenarios, it is necessary to perform a refined classification of these scenarios. This paper extracts the multidimensional statistical features of the scenarios, establishes a feature–application mapping relationship based on the practical application characteristics of power systems, and employs a quantile threshold method to classify extreme scenarios into the following three typical application scenarios:

(1) Continuous low output scenarios

The continuous low output of renewable energy is the key reason for the imbalance in the new power system. Generating such scenarios will provide data support for planning model safety constraint verification and reserve capacity analysis.

This study selects the lowest 30% of daily average values as the continuous low output scenario set.

(2) Short-term high ramp rate and high fluctuation frequency scenarios

These scenarios depict sudden surges or drops in wind power output over short timeframes (15 min intervals), which can trigger grid frequency instability, voltage excursions, and even cascade failures such as load shedding or unit uncoupling. Generating such scenarios aids in optimizing safety defense systems and validating control strategy effectiveness.

This study selected the top 30% of scenarios with the maximum ramp rates (15 min intervals) and fluctuation frequencies as the short-term high ramp rate and high fluctuation frequency scenario set.

(3) Medium and long-term high ramp rate and peak–valley difference scenarios

Unlike short-term scales, this scenario extends ramp events to 30 min intervals or longer, focusing on describing significant output fluctuations and pronounced peak–valley differences over medium and long-term timeframes. The generated results for these scenarios provide data support for resource optimization and medium and long-term reserve capacity planning.

This study selected the top 30% of scenarios ranked by maximum ramp rate (30 min time interval) and peak–valley difference as the medium and long-term high ramp rate and peak–valley difference scenario set.

Significant diversity still exists within these application-defined categories. Direct modeling of these scenarios would struggle to capture their consistent probability distribution patterns. To enhance the accuracy and plausibility of generated scenarios, each category requires clustering analysis prior to generation to identify subcategories with similar variation patterns. Subsequent extreme scenario generation proceeds based on these

subcategory classifications. Specifically, this study employs the K -means clustering algorithm based on Euclidean distance, with the number of clusters K determined using the silhouette coefficient method.

2.3. Framework for Generating Extreme Wind Power Output Scenarios

Due to the scarcity of sample data under extreme weather conditions, generating representative extreme output scenarios has become a critical step in enhancing the new power system's ability to respond to extreme events. To this end, this paper proposes a method for generating extreme wind power output scenarios guided and constrained by statistical features, based on the typical statistical features of extreme wind power output scenarios. The overall framework is shown in Figure 2.

Step 1: Extreme scenario definition and screening. Four key statistical features are selected: daily average, peak–valley difference, maximum ramp rate, and fluctuation frequency. Based on their distinct physical meanings, daily average and peak–valley difference are categorized as global static features to assess overall output levels, while maximum ramp rate and fluctuation frequency are classified as local dynamic features to capture abrupt temporal variations. A comprehensive evaluation system integrating these four-dimensional features is then constructed to identify and select extreme scenario samples from historical data.

Step 2: Application-based classification. By analyzing differentiated demands for extreme output features across various power system application scenarios, multiple categories of extreme scenario sets were established.

Step 3: Cluster analysis. The K -means algorithm was employed to cluster scenarios, aiming to identify subcategories with similar variation patterns and provide a data foundation for subsequent scenario generation.

Step 4: Scenario generation. Scenarios are generated within each application subset. To enhance generation quality, the WGAN-GP model (based on Wasserstein distance and gradient penalty) is employed to improve training stability of the generative adversarial network. Statistical feature guidance mechanisms and statistical feature loss function constraints are introduced into the generator.

Step 5: Scenario evaluation. A comprehensive assessment is conducted based on both accuracy and diversity of generated scenarios.

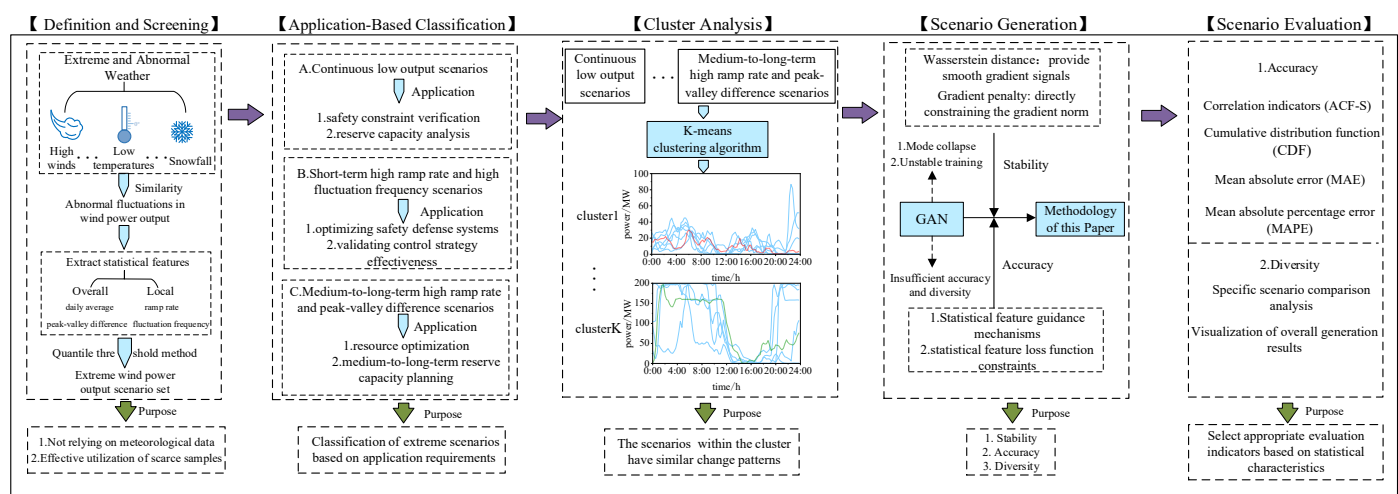


Figure 2. A framework for generating extreme wind power output scenario guided and constrained by statistical features.

2.4. Generative Adversarial Network

GAN consists of two components: a generator (G) and a discriminator (D). Its core idea lies in approximating the Nash equilibrium through adversarial training between the generator and discriminator. The objective function is shown in Equation (5) [14]:

$$\min_G \max_D V(D, G) = \mathbb{E}_{x \sim P_{data}(x)} [\log D(x)] + \mathbb{E}_{z \sim P_z(z)} [\log(1 - D(G(z)))] \quad (5)$$

In the equation, $D(x)$ represents the probability that the discriminator classifies the real sample x as real data; $D(G(z))$ represents the probability that the discriminator classifies the generated sample $G(z)$ as real data; and z is the random noise vector input to the generator. By optimizing this adversarial loss function, the generator G can gradually generate samples that approximate the distribution of real data, while the discriminator D continuously improves its ability to distinguish between real and generated samples.

However, extreme wind power output scenarios generally have the characteristics of “insufficient samples, extreme fluctuations, and skewed distributions.” When using traditional GAN networks for modeling, key challenges emerge, including gradient instability and mode collapse due to sample scarcity and extreme fluctuations, as well as evaluation metric failure caused by JS divergence saturation under highly skewed distributions.

2.5. Improvement of WGAN-GP Based on Wasserstein Distance and Gradient Penalty

To address these challenges, we construct a WGAN-GP model. The Wasserstein distance replaces JS divergence to provide stable gradients for learning complex distributions. The gradient penalty is then used to enforce the Lipschitz constraint more flexibly than weight clipping, ensuring training stability while maximizing the model's expressive capacity.

The loss function of the WGAN-GP model consists of two parts: the generator loss and the evaluator loss, as shown in Equations (6) and (7) respectively [14]. The optimization objective of the generator is to maximize the evaluator's score for the generated samples; the evaluator needs to maximize the score for the real samples and minimize the score for the generated samples, and introduces a gradient penalty term in the loss function to enhance the stability of the evaluator's gradient and satisfy the Lipschitz constraint.

$$L_G = -\mathbb{E}_{\tilde{x} \sim P_g} [D(\tilde{x})] \quad (6)$$

$$L_D = \mathbb{E}_{\tilde{x} \sim P_g} [D(\tilde{x})] - \mathbb{E}_{x \sim P_r} [D(x)] + \lambda \mathbb{E}_{\hat{x} \sim P_{\hat{x}}} \left[\left(\|\nabla_{\hat{x}} D(\hat{x})\|_2 - 1 \right)^2 \right] \quad (7)$$

In the equation, P_r and P_g respectively represent the probability distributions of the real samples and the generated samples; $D(\tilde{x})$ and $D(x)$ are the score outputs of the evaluator for the real samples and the generated samples; and $\hat{x}$ is the sampling point obtained by randomly interpolating between the real samples and the generated samples.

2.6. Statistical Feature Guidance and Constraint Mechanism

The WGAN-GP model designed for extreme wind power output scenarios enhances training stability through the gradient penalty mechanism and performs exceptionally well in generating diverse scenarios. However, extreme wind power output scenarios exhibit significant non-Gaussian characteristics and strong volatility. Only relying on adversarial training with Wasserstein distance constraints is insufficient to ensure the consistency of generated scenarios with the real extreme scenarios in key statistical features.

To address these issues, this paper adopts a statistical feature guidance and constraint mechanism.

2.6.1. Generation of Statistical Feature Vectors

To embed key statistical features into the generation process and guide the generation of scenarios, it is first necessary to construct a feature sampler that can represent the statistical features of historical extreme scenarios. Due to the scarcity of historical data for extreme scenarios, using complex parametric distribution models (such as Gaussian mixture models) is prone to overfitting and unstable training. To address this, this paper adopts a sampling strategy based on weighted averaging, directly learning and generating new feature vectors from historical features, with the core being to maintain the distribution characteristics of the feature space.

For the given set of real extreme wind power output scenarios, the four statistical features are calculated to form a multidimensional feature vector set $V = \{v_1, v_2, \dots, v_N\}$, where $v_i \in \mathbb{R}^4$ and N represent the number of scenarios.

The new feature vector is generated through the weighted average of historical feature vectors, and its mathematical expression is as follows:

$$\hat{v} = \lambda v_i + (1 - \lambda) v_j \quad (8)$$

In the equation, $\hat{v}$ represents the newly generated feature vector; v_i and v_j are two different basic feature vectors randomly selected from the historical feature set; λ is the combination weight coefficient, which follows $\lambda \sim U(0,1)$.

To ensure the consistency of all features on the numerical scale, all features are normalized to the range of $[-1, 1]$. This weighted averaging operation ensures that the newly generated feature vector lies within the space spanned by the historical feature vectors, thereby effectively avoiding the generation of physically unreasonable or overly outlier feature points.

2.6.2. Statistical Feature Self-attention Guidance

After obtaining the feature vectors that can represent the statistical features of extreme scenarios, in order to flexibly integrate them into the internal process of the generation model, this paper designs a statistical feature-guided self-attention mechanism module. This module injects the statistical features as bias terms into the attention calculation, guiding the model to pay more attention to the temporal dependencies that are consistent with the historical statistical features of extreme scenarios during the generation process. The specific steps are as follows:

(1) Statistical feature embedding

Firstly, a batch of statistical feature vectors $F \in \mathbb{R}^{B \times 4}$ is sampled, where B represents the batch size. These vectors are mapped to a high-dimensional space through a feature encoding network to obtain a bias vector D , as shown in Equation (9):

$$D = W_2 \times \text{LeakyReLU}(W_1 \times F + b_1) + b_2 \quad (9)$$

In the equation, $W_1 \in \mathbb{R}^{B \times 4}$ and $W_2 \in \mathbb{R}^{d_h \times B}$ are learnable parameters, while d_h represents the hidden dimension.

(2) Bias term generation

The bias vector D is split into query bias $d_k \in \mathbb{R}^{B \times d_k}$ and key bias $d_q \in \mathbb{R}^{B \times d_k}$, where d_k represents the dimension of the attention key query. This split design enables

the model to independently adjust the query vector and key vector of the attention mechanism.

(3) Attention calculation

Given input feature $X \in \mathbb{R}^{B \times C \times L}$, the basic query Q_0 , key K_0 and value V are obtained through convolution projection.

After injecting statistical feature bias, the final query matrix Q and key matrix K are formed, as shown in Equations (10) and (11):

$$Q = Q_0 + d_q \quad (10)$$

$$K = K_0 + d_k \quad (11)$$

Finally, the attention weight calculation is carried out, as shown in Equation (12):

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (12)$$

(4) Residual connection

The final output is connected to the input residuals through the learnable parameter γ , as shown in Equation (13); while retaining the original feature information, the attention mechanism is allowed to dynamically adjust the temporal dependency according to the statistical features.

$$X_{out} = \gamma \times \text{Attention}(Q, K, V) + X \quad (13)$$

2.6.3. Statistical Feature Loss Function Constraint

In the training of the generator, while considering the adversarial loss, an additional statistical feature loss function is introduced to constrain the consistency of the generated scenarios and the real scenarios in key statistical features.

The total loss function of the generator is shown in Equation (14) [14]:

$$L_G = -\mathbb{E}_{\tilde{x} \sim P_g} [D(\tilde{x})] + \alpha \times L_{stats} \quad (14)$$

In the equation, $-\mathbb{E}_{\tilde{x} \sim P_g} [D(\tilde{x})]$ represents the adversarial loss; L_{stats} represents the statistical feature loss; α represents the weight of the statistical feature loss.

In this paper, the daily average, peak–valley difference, maximum ramp rate and fluctuation frequency of the extreme wind power output scenarios are taken as the key statistical features, and a statistical feature loss function L_{stats} is constructed based on them, as shown in Equation (15):

$$L_{stats} = \frac{1}{B} \sum_{i=1}^B \sum_{j=1}^F W_j \left| \phi_j(X_g^{(i)}) - \phi_j(X_r^{(i)}) \right| \quad (15)$$

In the equation, F represents the set of statistical features; W_j represents the corresponding weights of the features; $\phi_j(\cdot)$ is the function for calculating the j th statistical feature; $X_g^{(i)}$ and $X_r^{(i)}$ are the corresponding real and generated samples in the batch.

2.6.4. WGAN-GP Model Guided and Constrained by Statistical Features

The statistical feature guidance and constraint mechanism proposed in this study operates through two synergistic components integrated into the WGAN-GP framework;

the overall framework is shown in Figure 3. First, the feature-guided attention module constrains the generation process by directing the generator's focus toward key statistical features during training. Second, the weighted feature loss function performs generation correction, explicitly shaping the output distribution to match targeted statistical features. Together, these mechanisms ensure that the generated scenarios not only approximate the overall data distribution but also accurately capture the specific extreme features for different applications.

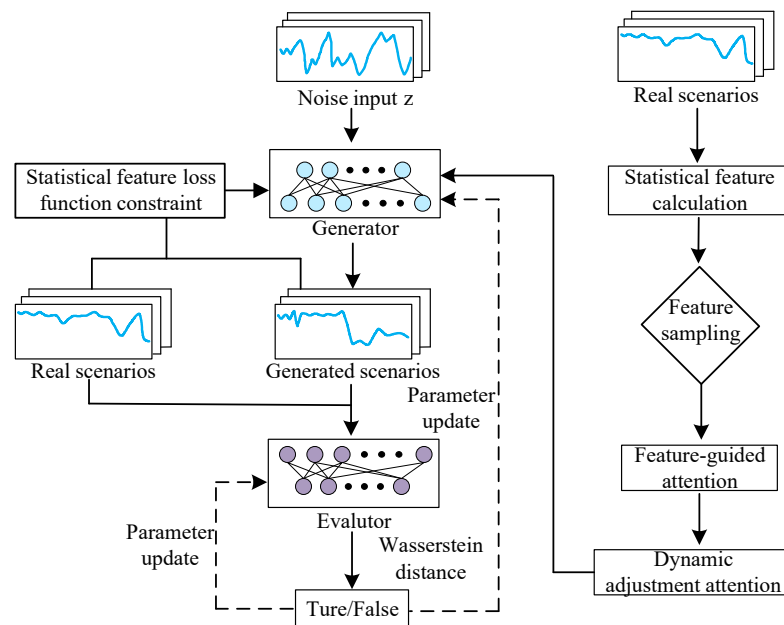


Figure 3. WGAN-GP model architecture guided and constrained by statistical feature.

3. Example Analysis

3.1. Dataset and Experimental Setup

The experiment utilized the wind power output data of a certain wind farm in 2019, with an installed capacity of 200 MW. The proportion of the training set and test set in the subsequent experiment is 60% and 40% respectively. The data time resolution was 15 min. After outlier elimination and missing value interpolation preprocessing, the experiment was conducted based on the Pytorch deep learning framework of Python 3.9 version. Under this experimental setup, the training time for the proposed method was approximately 50 min. In contrast, the conventional VAE and GAN models completed training much faster, and the WGAN-GP required about 20 min.

This paper implements the generation of extreme wind power output scenarios based on the WGAN-GP framework guided and constrained by statistical features. The specific training settings are as follows: the total number of iterations is 5000; the batch size is dynamically set according to the number of samples in each application cluster; to stabilize the training, the generator is updated once every two updates of the evaluator. The network structure is shown in Table 1, where K , S , P , B represent the size of the convolution kernel, stride, padding and batch size, and Conv represents the convolution layer.

Table 1. Network structure of WGAN-GP.

Network Composition	Construction	Parameter/Dimension	Activation Function	Hyperparameter
Generator	Input layer	Linear	(100, 256)	LeakyReLU
	Upsampling layer	Interpolate	(256, 96)	-
	Sequence generation layer	Conv1d	(1, 64), K5, S1, P2	LeakyReLU
		FeatureGuidedAttention	(64, 64)	-
		Conv1d	(64, 128), K5, S1, P2	LeakyReLU
		FeatureGuidedAttention	(128, 128)	-
Evaluator	Output layer	Conv1d	(128, 1), K5, S1, P2	Tanh
	Input layer	Unsqueeze	(B, 1, 96)	-
	Feature extraction layer	Conv1d	(1, 64), K5, S1, P2	LeakyReLU
		Conv1d	(64, 128), K5, S1, P2	LeakyReLU
	Output layer	Linear	(128 × 96, 1024)	LeakyReLU
		Linear	(1024, 1)	-

3.2. Evaluation Indicator

To analyze the accuracy and diversity of the generated scenarios, this paper uses the mean absolute error (MAE) to measure the reconstruction error between the real scenarios and the generated scenarios. The Pearson correlation coefficient between the autocorrelation function (ACF) sequences of the real scenarios and the generated scenarios is calculated as the indicator for evaluating the similarity of temporal dependency structure, and is denoted as ACF-S. The similarity of the overall distribution between the real scenarios and the generated scenarios is compared through the cumulative distribution function (CDF). In addition, starting from four statistical features, the mean absolute percentage error (MAPE) (applicable to peak–valley difference and maximum ramp rate) and the mean absolute error (applicable to daily average) are calculated to comprehensively evaluate the statistical features retention ability of the generated scenarios, as shown in the following equations:

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^N \left(\frac{1}{M} \sum_{d=1}^M |X_{\text{real}(i)} - X_{\text{gen}(d)}| \right) \quad (16)$$

$$\text{MAPE} = \frac{1}{N} \sum_{i=1}^N \left(\frac{1}{M} \sum_{d=1}^M \frac{|X_{\text{real}(i)} - X_{\text{gen}(d)}|}{|X_{\text{real}(i)}|} \right) \times 100\% \quad (17)$$

$$\text{ACF-S} = \frac{1}{N} \sum_{i=1}^N \left(\frac{1}{M} \sum_{d=1}^M \rho(\text{ACF}(X_{\text{real}(i)}), \text{ACF}(X_{\text{gen}(d)})) \right) \quad (18)$$

In the equation, N represents the total number of real scenarios, M represents the total number of generated scenarios that match each real scenario, X_{real} is the sequence of real scenarios, X_{gen} is the sequence of generated scenarios, and ρ is the calculation of the Pearson correlation coefficient.

3.3. Definition and Selection of Extreme Wind Power Output Scenarios

To screen extreme wind power output scenarios, this study selected four statistical indicators: daily average, peak–valley difference, maximum ramp rate, and fluctuation

frequency. These indicators were used to characterize the extreme features from both the overall distribution and local dynamic perspectives. For the statistical indicator, the quantile method was employed to set specific thresholds (0.1, 0.2, 0.3, 0.8, and 0.9), which were chosen to effectively capture extreme distribution tails while retaining a sufficient number of scenarios for subsequent analysis, and the scenarios were selected by combining the cutoff conditions of different thresholds.

(1) Overall distribution

Joint screening of daily average and peak–valley differences. The daily average reflects the overall power output level and is used to identify sustained high or low power output events (such as extreme high winds or calm days); the peak–valley difference quantifies the range of extreme daily fluctuations and screens out scenarios with significant power fluctuations.

Figure 4 shows the joint screening results based on daily average and peak–valley differences. Multiple quantile threshold lines are plotted in the figure to divide the entire plane into multiple regions, such as 0.1, 0.2, 0.8, and 0.9, and screen out extreme scenarios with specific power features.

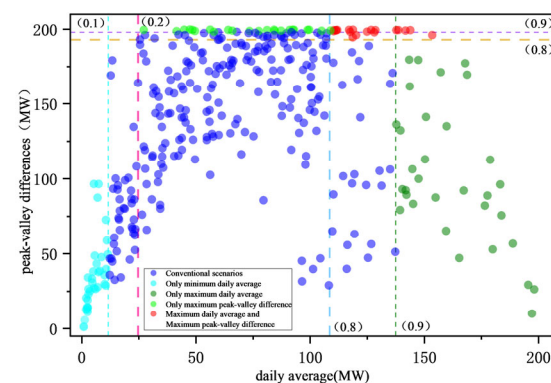


Figure 4. Overall statistical features screening results.

(2) Local dynamic

Maximum ramp rate and fluctuation frequency are jointly screened. The maximum ramp rate reflects the instantaneous sharp change in power; the fluctuation frequency quantifies the switching frequency of power direction within a day, reflecting the repetitive fluctuation characteristics of power.

Figure 5 shows the joint screening results based on fluctuation frequency and maximum ramp rate. Similarly, multiple quantile threshold lines are plotted in the figure to divide the entire plane into multiple regions, such as 0.8 and 0.9, and screen out extreme scenarios with specific power features.

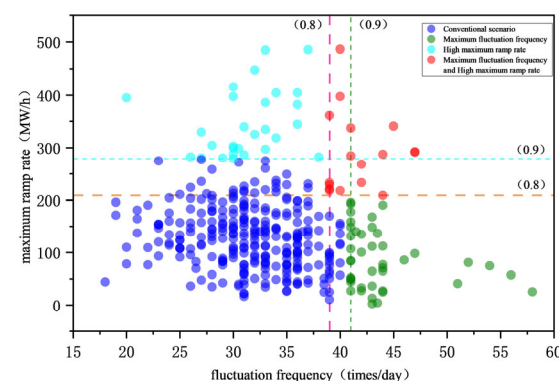


Figure 5. Localized statistical features screening results.

3.4. Classification of Extreme Scenarios in Application Requirements

Based on the comprehensive extreme wind power output scenario set, the quantile threshold method is used to establish three types of typical application scenarios. Further, each type of typical application scenario is clustered to achieve a refined classification of extreme scenarios. The clustering results are shown in Table 2.

Table 2. Typical application clustering results.

Application	Optimal Number of Clusters	Number of Scenarios Within a Cluster
Application 1	4	7, 32, 5, 4
Application 2	2	8, 7
Application 3	4	7, 6, 6, 4

(1) Application 1—continuous low output scenario

The continuous low output scenario set is shown in Figure 6. Each clustered typical scenario in the figure represents the cluster center.

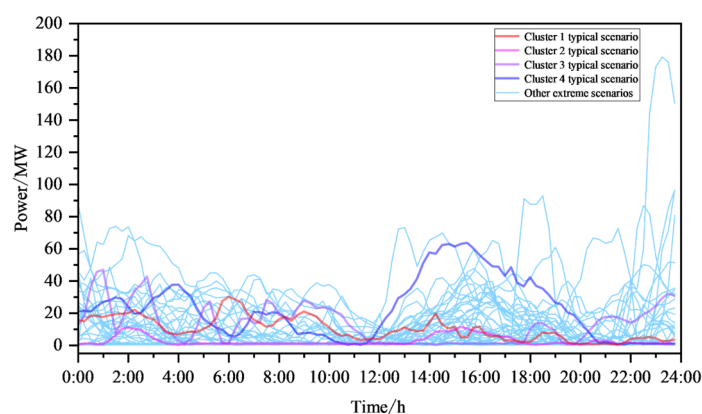


Figure 6. Continuous low output extreme scenario set.

(2) Application 2—Short-term high ramp rate and high fluctuation frequency scenario

The short-term high ramp rate and high fluctuation frequency scenario set are shown in Figure 7.

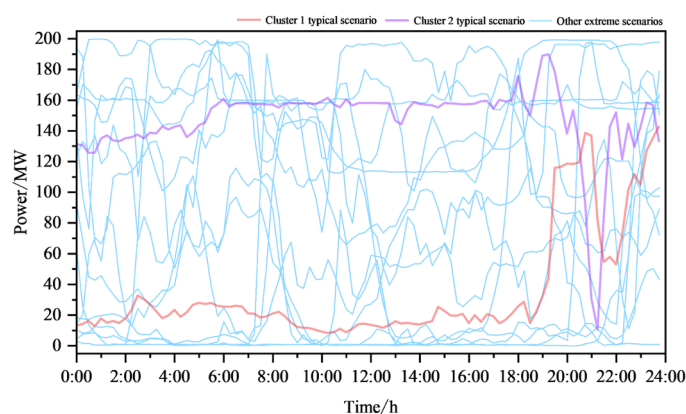


Figure 7. Extreme scenario set with short-term high ramp rate and high fluctuation frequency.

(3) Application 3—Medium and long-term high ramp rate and peak–valley difference scenario

The medium and long-term high ramp rate and high peak–valley difference scenario sets are shown in Figure 8.

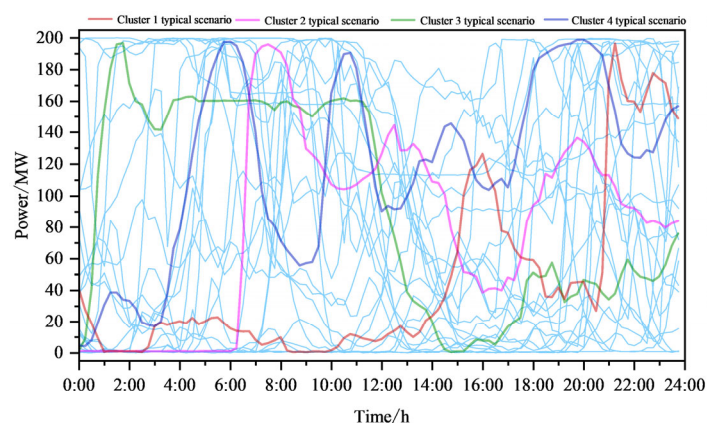


Figure 8. Extreme scenario set with medium and long-term high ramp rate and high peak–valley difference.

3.5. Generation and Evaluation of Extreme Scenarios

This paper analyzes three typical application scenarios. The comparison models include traditional VAE, GAN, WGAN-GP, and the method proposed in this paper. The evaluation content is carried out from two dimensions: accuracy and diversity. In terms of accuracy, it is quantitatively evaluated through mean absolute error (MAE), autocorrelation function similarity (ACF-S), cumulative distribution function (CDF), and mean absolute percentage error (MAPE) of key statistical features; in terms of diversity, it is comprehensively evaluated by comparing the generated images with the real images, to assess the generation quality of different models in various application scenarios.

3.5.1. Comparison of Continuous Low-Output Scenario Generation

The study analyzes the accuracy of the generated scenarios produced by each generative model under different clusters. Table 3 presents the evaluation indicator results of the generated scenarios produced by different generative models under each cluster. Figure 9 further shows the CDF curves of wind power output for the generated scenarios and the real scenarios by each method. To analyze the diversity of the generated scenarios, Figure 10 compares the overall generated scenarios against the real ones, alongside a presentation of their local details.

Table 3. Evaluation indicators for generated scenarios under different methods of application 1.

Cluster	Method	MAE/MW	ACF-S	Daily Average MAE/MW
Cluster 1	VAE	4.50	0.83	1.42
	GAN	4.11	0.84	1.34
	WGAN-GP	4.16	0.90	1.10
	The proposed method	2.72	0.96	0.74
Cluster 2	VAE	4.02	0.82	1.92
	GAN	3.83	0.85	1.82
	WGAN-GP	3.65	0.84	1.61
	The proposed method	3.39	0.86	1.42
Cluster 3	VAE	4.46	0.92	2.45
	GAN	4.31	0.95	2.05
	WGAN-GP	2.89	0.96	1.00
	The proposed method	2.71	0.98	0.73
Cluster 4	VAE	7.51	0.92	3.11
	GAN	6.80	0.95	2.88

WGAN-GP	5.49	0.94	1.23
The proposed method	4.34	0.96	1.19

As shown in Table 3, within each cluster, the proposed method outperforms VAE, GAN and WGAN-GP models in accuracy for extreme generated scenarios. The baseline VAE model highlights its limitation in capturing complex, extreme patterns compared to adversarial learning frameworks. Among the GAN variants, the introduction of Wasserstein distance and gradient penalty (WGAN-GP) over the original GAN leads to more stable training improvements. Most importantly, our proposed method—built on the WGAN-GP framework and further enhanced with statistical feature guidance and constraints—demonstrates an improvement in accuracy. Specifically, the minimum MAE value is 2.71, and the maximum difference compared to other models is 2.46 (cluster 4); the maximum ACF-S value is 0.98, and the maximum difference compared to other models is 0.12 (cluster 1); the minimum daily average MAE value is 0.73, and the maximum difference compared to other models is 1.69 (cluster 4). The improvement achieved by the proposed method varies across clusters, which can be attributed to their complicated data structures and distributional characteristics. For example, Cluster 2 contained a larger number of real scenarios with more intricate internal patterns, resulting in lower evaluation indicator quality compared to other clusters. Meanwhile, the MAE indicator shows more substantial improvement compared to ACF-S. This is because the feature loss function directly constrains these statistical features. As shown in Figure 9, comparing the CDF curves of scenarios generated by different methods with real scenarios reveals that the scenarios generated by the proposed method (solid red line) exhibit a higher degree of fit with real scenarios (dashed black line). Furthermore, the Wasserstein distances between the generated scenarios and real scenarios, indicated in parentheses in the legend, provide quantitative support for this conclusion.

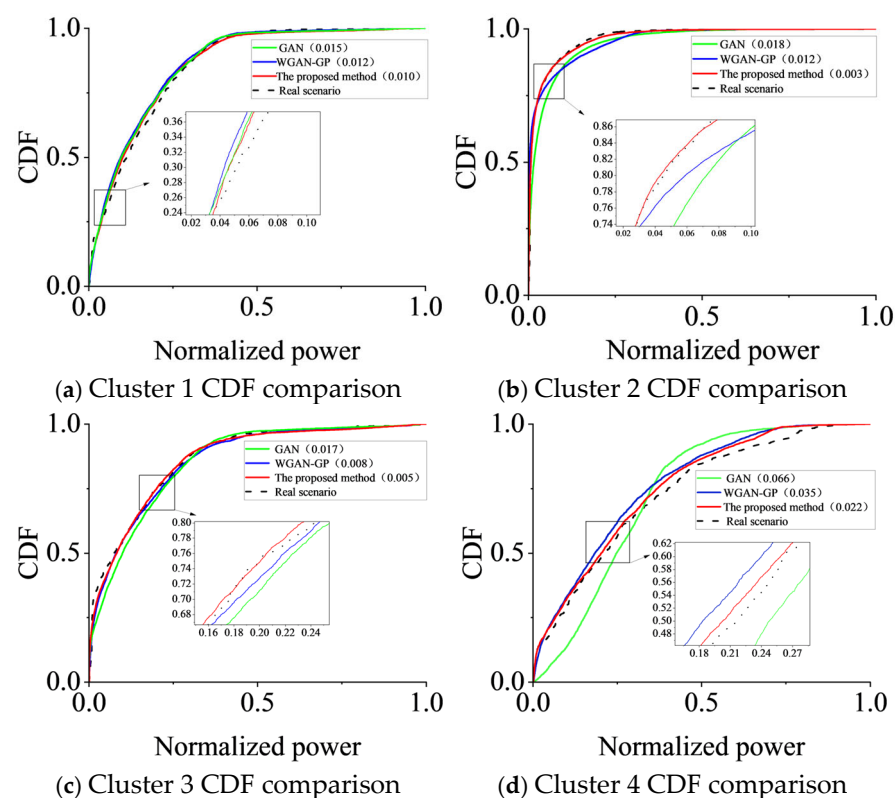


Figure 9. CDF comparisons of application 1.

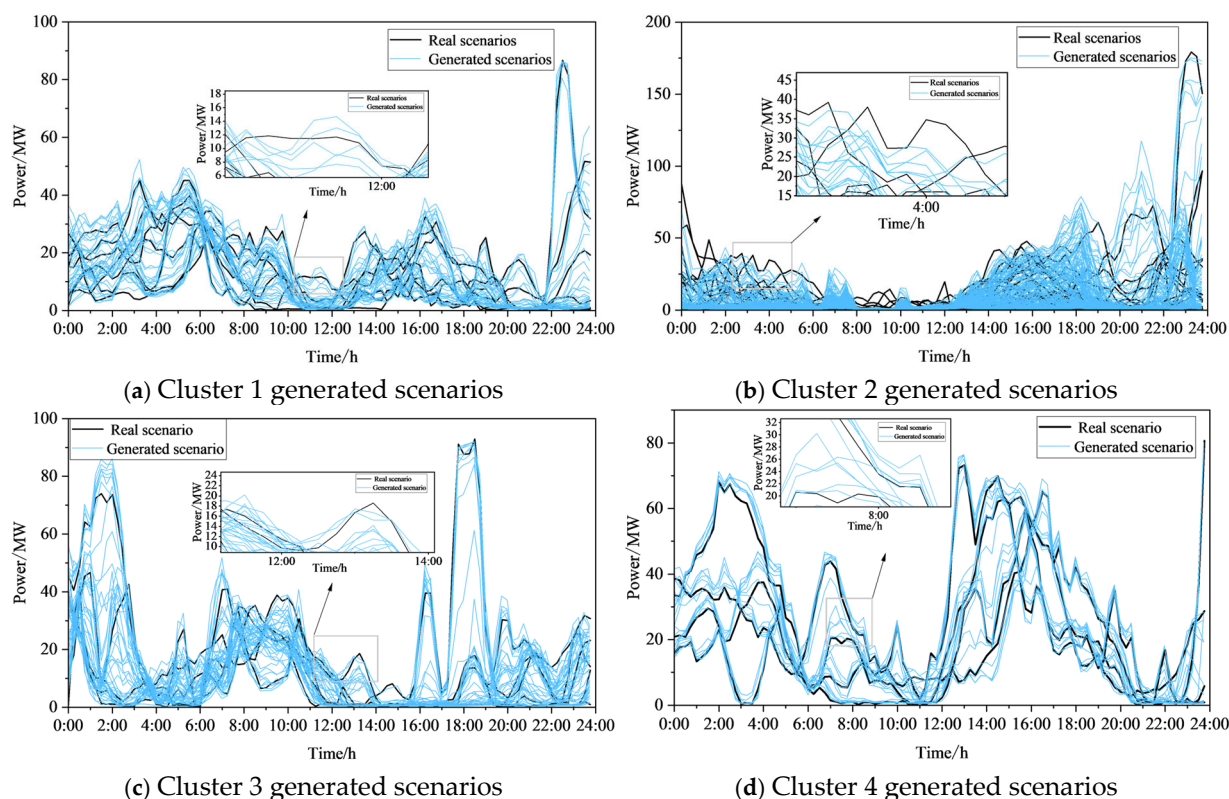


Figure 10. Results generated by continuous low-output extreme scenario.

The data results in Table 3 and Figure 9 demonstrate that the proposed method has a significant advantage in the accuracy of generating the statistical features of the scenarios. In terms of the diversity of generated scenarios, Figure 10 shows the generated samples under different clusters. As shown in Figure 10a, which represents the generated scenarios under Cluster 1, the results indicate that our method can effectively capture and reproduce reasonable detail fluctuations and individual differences in extreme scenarios while maintaining the accuracy of key statistical features.

3.5.2. Comparison of Short-Term High Ramp Rate and High Fluctuation Frequency Scenario Generation

The study analyzes the accuracy of the generated scenarios produced by each generative model under different clusters. Table 4 presents the evaluation indicator results of the generated scenarios produced by different generative models under each cluster. Figure 11 further compares the CDF curves of the generated scenarios and the real scenarios by each method. To analyze the diversity of the generated scenarios, Figure 12 plots the typical scenarios and the overall generated results under each cluster.

Table 4. Evaluation indicators for generated scenarios under different methods of application 2.

Cluster	Method	ACF-S	Ramp Rate MAPE	Fluctuation Frequency MAPE
Cluster 1	VAE	0.90	26.85%	19.55%
	GAN	0.89	25.79%	18.77%
	WGAN-GP	0.94	20.67%	14.19%
	The proposed method	0.95	19.23%	12.82%
Cluster 2	VAE	0.81	26.67%	19.98%
	GAN	0.77	25.40%	19.15%
	WGAN-GP	0.93	19.80%	12.10%

The proposed method	0.94	18.04%	11.39%
---------------------	------	--------	--------

As shown in Table 4, within each cluster, the WGAN-GP model significantly improves the ACF-S values compared with the original VAE and GAN model, while the MAPE values of the statistical features generally decrease, indicating that the introduction of the Wasserstein distance and the gradient penalty mechanism effectively enhances the stability of the training process compared to the original GAN, thereby improving the overall quality of the generated scenarios. The proposed method further incorporates statistical feature guidance and constraint mechanisms on the basis of WGAN-GP, which continues to increase the ACF-S values and significantly reduces the MAPE values of the statistical features. The improvement in ramp rate MAPE and fluctuation frequency MAPE is substantially greater than that in ACF-S. Figure 11 shows that the proposed method (solid red line) achieves the closest fit to the real scenarios (dashed black line) and the best distribution consistency, as evidenced by the CDF comparison and Wasserstein distances. Combined with the results in Table 4, this demonstrates a clear advantage for the proposed method in accurately replicating the statistical features of the scenarios.

Regarding diversity, Figure 12 presents generated scenario samples under different clusters. For instance, Figure 12a demonstrates that the overall fluctuation trend of the real scenarios is maintained (left), while the detailed view (right) shows that a single real scenario corresponds to multiple varied yet plausible scenarios. These results indicate that the proposed method not only preserves statistical accuracy but also effectively captures the inherent uncertainty of extreme scenarios, thus producing a realistic and diverse set of scenarios.

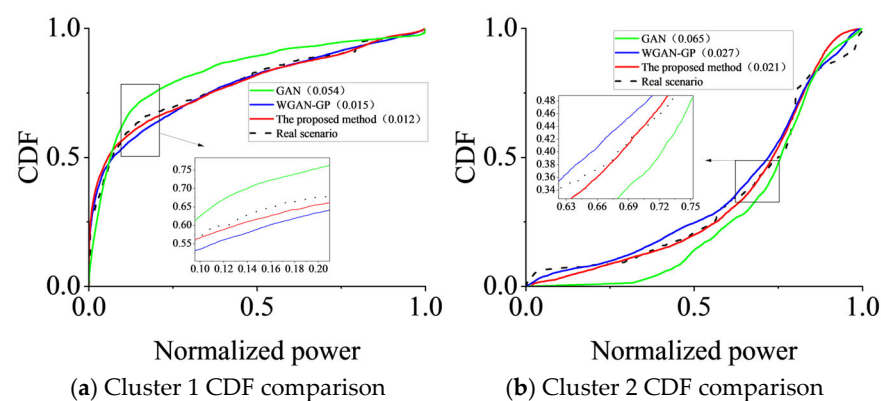
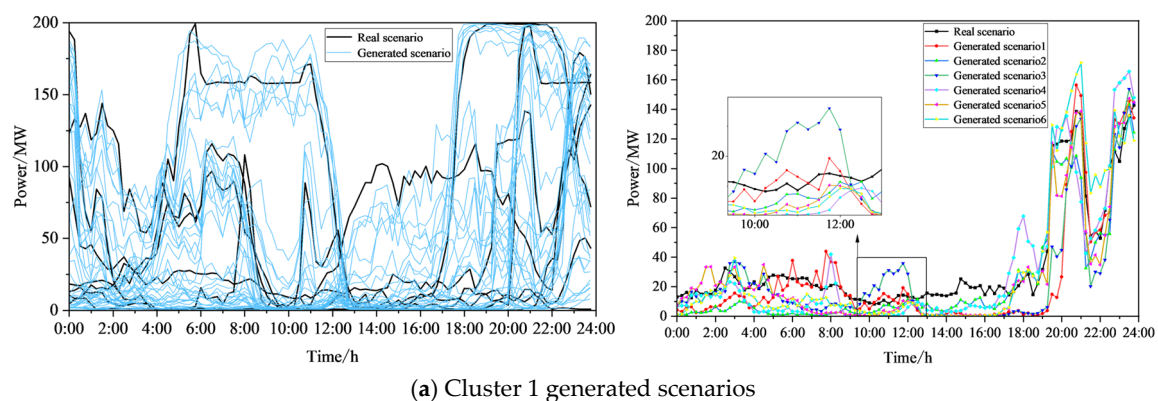


Figure 11. CDF comparisons of application 2.



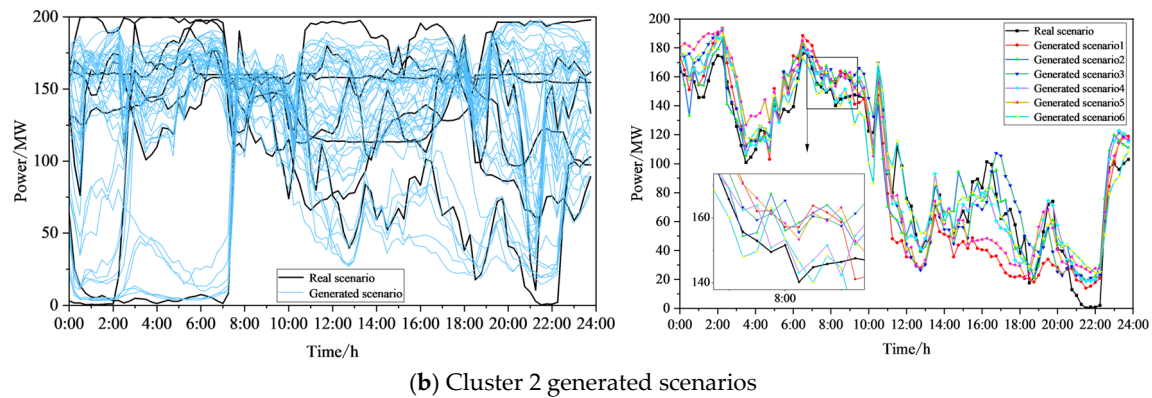


Figure 12. Results generated by short-term high ramp rate and high fluctuation frequency extreme scenario.

3.5.3. Comparison of Medium and Long-Term High Ramp Rate and Peak–Valley Difference Scenario Generation

As shown in Table 5, the proposed method achieves optimal performance across all clusters in application 3. The specific comparisons are as follows: First, the original GAN suffers from mode collapse, resulting in lower ACF-S values and indicating poor generation quality. Its acceptable performance on certain individual indicators (such as peak–valley difference MAPE in clusters 1 and 3) further reveals a lack of diversity, as it merely repeats limited patterns, thereby limiting the reference value of its statistical results. Second, compared with the WGAN-GP model, the proposed method further improves the ACF-S values and significantly reduces the MAPE for ramp rate and peak–valley difference, demonstrating that the generated scenarios not only preserve the temporal structure but also enhance the reconstruction accuracy of key statistical features, validating the effectiveness of the statistical feature guidance and constraint mechanism.

Table 5. Evaluation indicators for generated scenarios under different methods of application 3.

Cluster	Method	ACF-S	Ramp Rate MAPE	Peak–Valley Difference MAPE
Cluster 1	VAE	0.89	26.25%	9.26%
	GAN	0.86	23.26%	3.76%
	WGAN-GP	0.94	21.24%	6.71%
	The proposed method	0.96	13.80%	4.95%
Cluster 2	VAE	0.88	32.56%	9.81%
	GAN	0.87	19.15%	5.47%
	WGAN-GP	0.96	29.18%	3.25%
	The proposed method	0.97	17.84%	2.99%
Cluster 3	VAE	0.92	27.49%	4.28%
	GAN	0.91	23.92%	2.00%
	WGAN-GP	0.96	25.22%	1.60%
	The proposed method	0.99	15.71%	1.58%
Cluster 4	VAE	0.81	26.12%	11.85%
	GAN	0.83	23.04%	9.14%
	WGAN-GP	0.96	15.27%	3.91%
	The proposed method	0.97	14.60%	2.51%

From the comparison of CDF curves and Wasserstein distances in Figure 13, it can be observed that the cumulative distribution of scenarios generated by the proposed method is closest to that of the real scenarios, indicating higher accuracy in the statistical

distribution. The results in Figure 14 show that the generated scenarios not only maintain the real fluctuation trend overall but also exhibit diverse detail variations. In summary, the proposed method demonstrates significant advantages in both statistical accuracy and diversity of the generated scenarios in application 3.

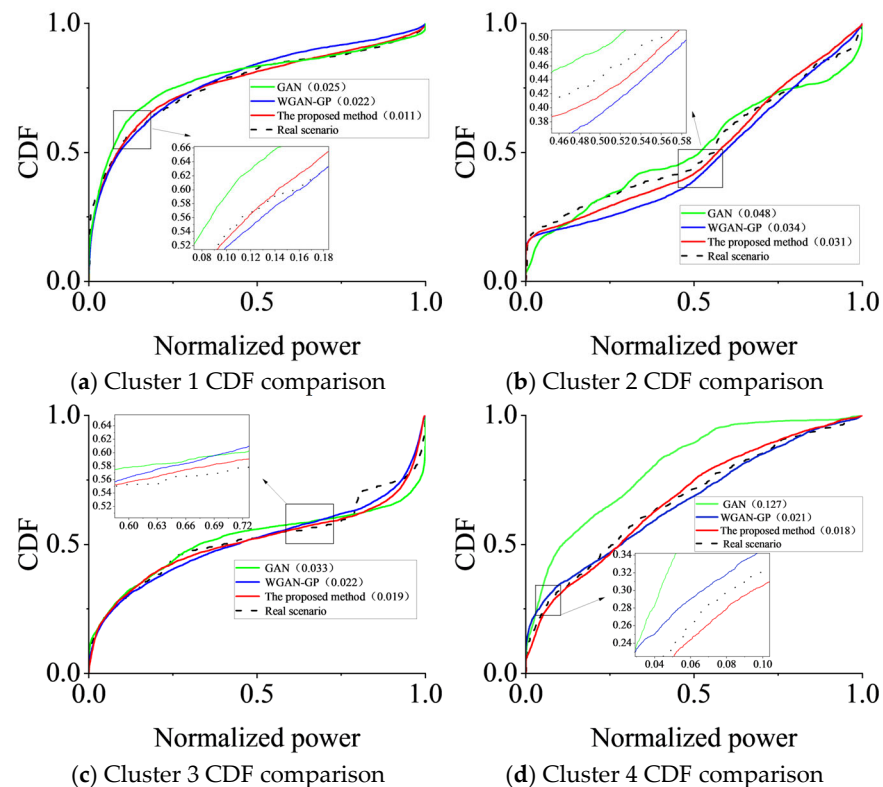
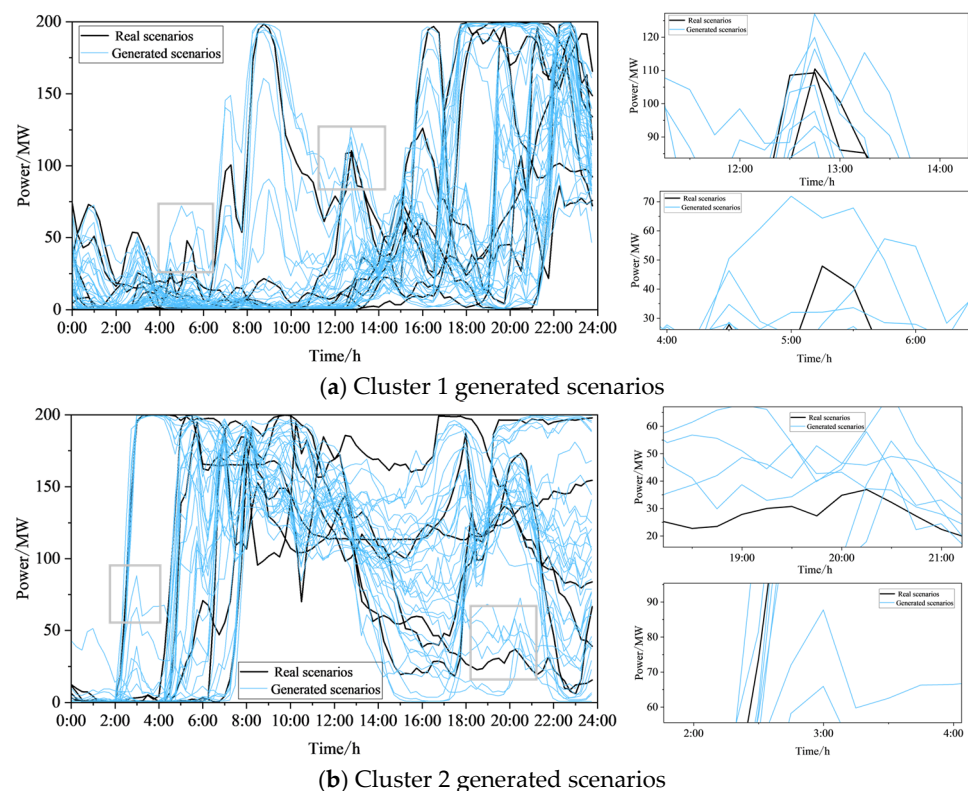


Figure 13. CDF comparisons of application 3.



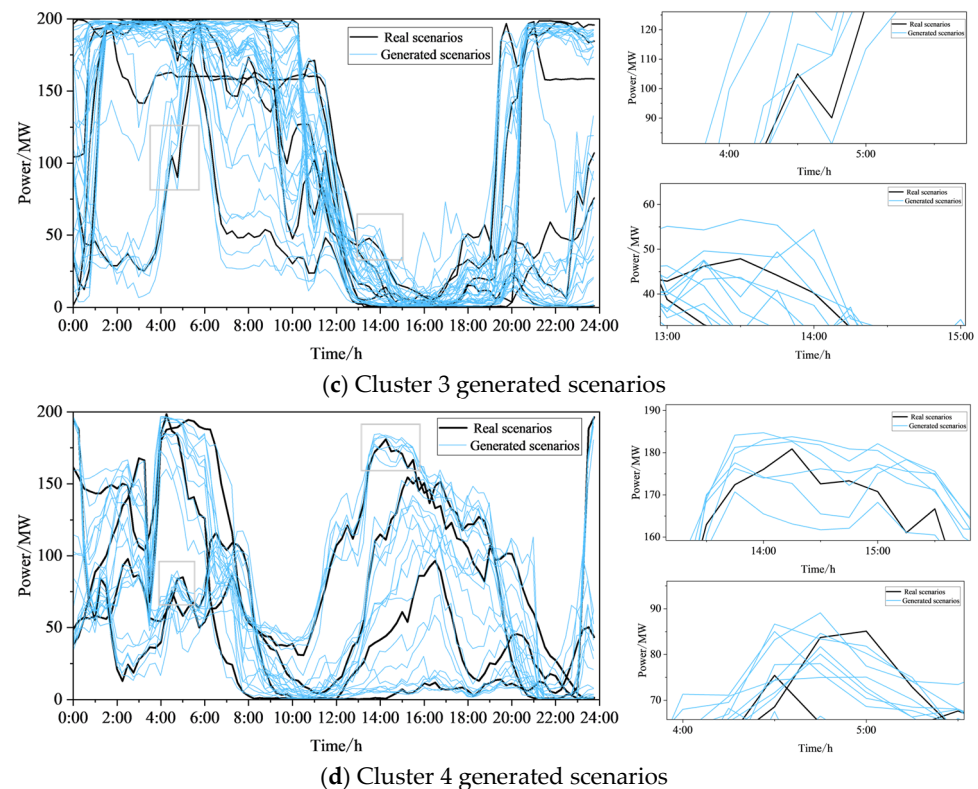


Figure 14. Results generated by medium and long-term high ramp rate and peak–valley difference extreme scenario.

4. Conclusions

This paper proposes a method for generating extreme wind power output scenarios guided by statistical features to address the scarcity of historical extreme data. The main conclusions are as follows:

- (1) The proposed screening framework, combining multidimensional statistical features and quantile thresholds, effectively extracts extreme patterns from limited data to provide quality samples for generation.
- (2) The refined clustering method for extreme scenarios generates tailored subsets for diverse applications. These generated scenarios exhibit strong relevance to specific power system applications, thereby enhancing their decision support value.
- (3) The improved GAN model, integrating Wasserstein distance, gradient penalty, statistical feature guidance and constraints, ensures stable training and generates scenarios that capture detailed fluctuations and individual differences under extreme conditions, but also maintains realistic fluctuations while maintaining high consistency with real data in statistical distribution.

Future research will focus on two directions: the first is exploring the generation of joint extreme scenarios for wind and solar power output; the second is the application of the generated extreme scenarios in stochastic unit commitment or economic dispatch models, aimed at enhancing the security and robustness of system decisions. Meanwhile, this work has several limitations that point to clear avenues for future investigation: the quality of generated scenarios relies on the quality of historical data and the chosen clustering structure. Furthermore, the current guidance relies primarily on statistical features; incorporating more explicit physical constraints (e.g., turbine characteristics, spatiotemporal correlations) could enhance the physical plausibility of the scenarios.

Author Contributions: Conceptualization, D.L.; data curation, B.Y.; methodology, D.L. and X.L.; resources, M.Q. and Y.Z.; software, X.L. and Z.L.; supervision, M.Q. and B.Y.; validation, X.L. and Y.Z.; visualization, Z.L.; writing—original draft, X.L.; writing—review & editing, D.L. and X.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the National Natural Science Foundation of China (Grant number 51807109).

Data Availability Statement: The original contributions presented in the study are included in the article; further inquiries can be directed to the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Song, Z.; Zhang, M.; Chi, Y.; Li, J.; Zheng, Y. Research on the coordinated optimization of energy storage and renewable energy in off-grid microgrids under new electric power systems. *Glob. Energy Interconnect.* **2025**, *8*, 213–224. <https://doi.org/10.1016/j.gloi.2024.12.004>.
2. Yu, X.; Wang, B.; Wang, W.; Guo, X.; Han, J.; Chen, X. Analysis of renewable resources in central China under the “double carbon” strategy. *Energy Rep.* **2022**, *8*, 361–373. <https://doi.org/10.1016/j.egyr.2022.09.147>.
3. Zheng, X.; Zhang, D.; Bie, Z.; Wu, X. Distributed robust planning for new power system considering uncertainty and frequency security. *Alex. Eng. J.* **2025**, *127*, 690–704. <https://doi.org/10.1016/j.aej.2025.04.087>.
4. Li, H.; Ren, Z.; Fan, M.; Li, W.; Xu, Y.; Jiang, Y.; Xia, W. A review of scenario analysis methods in planning and operation of modern power systems: Methodologies, applications, and challenges. *Electr. Power Syst. Res.* **2022**, *205*, 107722. <https://doi.org/10.1016/j.epsr.2021.107722>.
5. Shen, Y.; Jiang, G.; Xu, J.; Feng, P.; Guo, F.; Wei, M.; Ma, Y. Multi-Spatiotemporal Power Source Planning for New Power Systems Considering Extreme Weathers. *Processes* **2025**, *13*, 3385–3385. <https://doi.org/10.3390/pr13113385>.
6. Li, J.; Zhou, J.; Chen, B. Review of wind power scenario generation methods for optimal operation of renewable energy systems. *Appl. Energy* **2020**, *280*, 115992. <https://doi.org/10.1016/j.apenergy.2020.115992>.
7. Kailash Chand, S.; Archee, G.; Rohit, B. Wind Power Scenario Generation Considering Spatiotemporal Correlations: A Distribution Free Hybrid VARMA-Copula Approach. *Smart Grids Sustain. Energy* **2023**, *8*, 17. <https://doi.org/10.1007/s40866-023-00175-0>.
8. Li, Y.; Peng, X.; Zhang, Y. Forecasting methods for wind power scenarios of multiple wind farms based on spatio-temporal dependency structure. *Renew. Energy* **2022**, *201*, 950–960. <https://doi.org/10.1016/j.renene.2022.11.002>.
9. Deng, J.; Li, H.; Hu, J.; Liu, Z. A new wind speed scenario generation method based on spatiotemporal dependency structure. *Renew. Energy* **2021**, *163*, 1951–1962. <https://doi.org/10.1016/j.renene.2020.10.132>.
10. Zhu, J.; Ai, Q.; Chen, Y.; Wang, J. Single-Location and Multi-Locations Scenarios Generation for Wind Power Based On WGAN-GP. *J. Phys. Conf. Ser.* **2023**, *2452*, 012022. <https://doi.org/10.1088/1742-6596/2452/1/012022>.
11. Dong, W.; Chen, X.; Yang, Q. Data-driven scenario generation of renewable energy production based on controllable generative adversarial networks with interpretability. *Appl. Energy* **2022**, *308*, 118387. <https://doi.org/10.1016/j.apenergy.2021.118387>.
12. Zhang, Y.; Ai, Q.; Xiao, F.; Hao, R.; Lu, T. Typical wind power scenario generation for multiple wind farms using conditional improved Wasserstein generative adversarial network. *Int. J. Electr. Power Energy Syst.* **2020**, *114*, 105388. <https://doi.org/10.1016/j.ijepes.2019.105388>.
13. Hu, S.; Zhu, R.; Tang, B. PV power scenario generation method based on wasserstein generative adversarial networks with gradient penalty. *Proc. CSU-EPSA* **2022**, *34*, 109–115. <https://doi.org/10.19635/j.cnki.csu-epsa.000993>.
14. Mi, Y.; Lu, C.; Shen, J.; Yang, X.; Ge, L. Wind power extreme scenario generation based on conditional generative adversarial network. *High Volt. Eng.* **2023**, *49*, 2253–2263. <https://doi.org/10.13336/j.1003-6520.hve.20230459>.
15. Niu, K.; Liu, J.; Tang, Y.; Han, Y.; Hu, J.; Liu, X.; Song, X. A method for generating annual system operating scenarios considering extreme meteorological events. *Power Syst. Technol.* **2024**, *48*, 3992–4001. <https://doi.org/10.13335/j.1000-3673.pst.2024.1239>.
16. Yang, Y.; Sun, Y.; Mao, Z.; Dong, X. Wind power extreme scenario generation caused by weather processes on medium time scale. *Proc. CSEE* **2025**, *23*, 9088–9098. <https://doi.org/10.13334/j.0258-8013.pcsee.241340>.

17. Wang, Y.; Pang, P.; Qi, B.; Wang, X.; Zhao, Z. A two-stage optimal pre-scheduling strategy for power system inertia assessment and replenishment under extreme weather events. *Sustain. Comput. Inform. Syst.* **2025**, *45*, 101079. <https://doi.org/10.1016/j.suscom.2024.101079>.
18. Chang, G.; Shan, R.; Li, X.; Xue, G.; Ma, R.; Su, S.; Wu, P.; Ruan, C. Multidimensional assessment and distributional characteristics of low-wind-power events in high proportion renewable energy power systems. *Power Syst. Technol.* **2025**, *49*, 1369–1378. <https://doi.org/10.13335/j.1000-3673.pst.2024.1422>.
19. Liu, T.; Ma, S.; Shen, J.; Fang, H. Research on flexible collaborative planning methods for source-grid-load-storage in new power systems. *J. Phys. Conf. Ser.* **2025**, *3110*, 012028. <https://doi.org/10.1088/1742-6596/3110/1/012028>.
20. Li, D.; Liu, H.; Zhang, J.; Han, X.; Zhang, S. Multistage power generation planning approach for new power systems with high renewable energy integration considering flexibility constraints. *Int. J. Electr. Power Energy Syst.* **2025**, *169*, 110818. <https://doi.org/10.1016/j.ijepes.2025.110818>.
21. Xu, G.; Guo, Z. Resilience enhancement of distribution networks based on demand response under extreme scenarios. *IET Renew. Power Gener.* **2023**, *18*, 48–59. <https://doi.org/10.1049/rpg2.12895>.
22. Xu, J.; Wang, B.; Sun, Y.; Xu, Q.; Liu, J.; Cao, H.; Jiang, H.; Lei, R.; Shen, M. A day-ahead economic dispatch method considering extreme scenarios based on wind power uncertainty. *CSEE J. Power Energy Syst.* **2019**, *5*, 224–233. <https://doi.org/10.17775/CSEEJPES.2016.00620>.
23. Qi, Y.; Liu, Y. Finite Control of High Risk Wind Power Ramping. *Proc. CSEE* **2013**, *33*, 69–75. <https://doi.org/10.13334/j.0258-8013.pcsee.2013.13.023>.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.